| (51) International Patent Classification 6 : | | (11) International Publication Number: | **WO 99/17194** |
|---|---|---|---|
| G06F 9/46, 17/60 | **A1** | (43) International Publication Date: | 8 April 1999 (08.04.99) |

(21) International Application Number: PCT/GB98/02944

(22) International Filing Date: 1 October 1998 (01.10.98)

(30) Priority Data:
97307745.6    1 October 1997 (01.10.97)    EP

(71) Applicant *(for all designated States except US)*: BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PUROHIT, Bharat [GB/GB]; 23 Potter Street, Northwood Hills, Middlesex HA6 1QJ (GB). SHEPHERDSON, John, William [GB/GB]; 4 Abbey Road, Sudbury, Suffolk CO10 6LA (GB). JUDGE, Donald [GB/GB]; 78 Leopold Road, Ipswich, Suffolk IP4 4RN (GB). ODGERS, Brian, Robert [GB/GB]; 15 Kingsway Grove, Lurgan, County Armagh BT66 7TE (GB).

(74) Agents: MUSKER, David, Charles et al.; R.G.C. Jenkins & Co., 26 Caxton Street, London SW1H 0RJ (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: RESOURCE MANAGEMENT SYSTEM

(57) Abstract

A resource handling system (120) has an intelligent software agent–based layer which provides a quasi–centralised control. The system has an interface to a user community and receives resource requests (115). There are multiple agents which negotiate to provide a response. The response can be substantially in real–time, rather than via a planning and scheduling capability, while there is also management of over– and under–bidding by the agents.

1

# RESOURCE MANAGEMENT SYSTEM

The present invention relates to process management and particularly to resource allocation in carrying out processes.

5     Resource allocation can often present a complex problem in an environment in which many tasks need to be carried out by various resources and with varying levels of priority. In practice, this is further complicated by changes in the situation, such as incoming tasks of high urgency and equipment failure.

There are systems already known for resource allocation. For instance, 10  an example of resource allocation is the automation of a business process. Documents, information or tasks can be passed from one resource to another for action according to a set of procedural rules. There is a description of such a system, published by the Workflow Management Coalition under the title "Workflow Reference Model Specification" on the Internet at the Universal 15 Resource Location (URL) http:\\www.aiai.ed.ac.uk:80\WfMC\ (now http://www.aiim.org/wfmc). The Coalition has proposed a framework for the establishment of workflow standards. This framework includes five categories of interoperability and communication standards that will allow multiple workflow products to exist and interoperate within a user's environment. The Coalition has 20  also published a white paper entitled "The Work of the Coalition".

Workflow in this context is relevant to such areas as image management systems, document management systems, relational or object database systems and electronic mail systems.

One approach to automating workflow (specifically, job scheduling) is 25 disclosed in GB 2194086, in which computer calculates scheduling for completing projects with limited resources (e.g. people).

The article "Resource Management in Large Distributed Systems", Goscinski A et al, Operating systems Review (SIGOPS), vol. 24, No. 4. 1 October 1990, pages 7-25, discloses sharing computing resources (for example printers 30  and computation) by using a resource management centre which negotiates with a local network. It does not discuss workflow.

The article "Using Intelligent Agents to Manage Business Processes", Jennings et al, Proc First International Conference on the Practice Application of .

Intelligent Agents and Multi Agent Technology (PAAM 96), London, 1996, pages 345-360, discusses the use of negotiating agents to manage workflow. The agents comprise, for example, a management module which negotiates for services which are needed, and other modules (which may be at a hierarchically

5   lower level) which can provide services through service execution modules, and which negotiate with the management module.

The results of the negotiation is to set up a long term service life cycle agreement (SLA) constituting an agreement to do a certain amount of work per unit of time on agreed terms. These may be re-negotiated.

10  The foregoing are generally concerned with work flow management, which is a process of allocating streams of work between different resources; in other words, allocating rates at which work is to be done. Thus, reallocation of the resources is only required where the rates at which work is to be done are changed. Examples are call handling in a call service centre; correspondence

15  handling in a correspondence service centre; manufacturing and assembly processes and so on.

General discussion of the application of agents to network management is disclosed in "Applying the Agent Paradigm to Network Management", Davidson et al, BTTJ Vol. 16, No. 3, July 1998.

20  According to embodiments of the present invention there is provided a resource allocation system comprising computing means for executing: at least one first computer program module for determining an allocation matching one or more of a plurality of resource supplies to a resource requirement; at least two second computer program module for monitoring the state of a corresponding one

25  of said resource supplies and for communicating with said first program module.

It will be understood that the invention is also applicable to similar processes; for example, to the distribution of telecommunications bit streams between several different channels having varying channel capacities. For example, a broad bandwidth bit stream may be split between several ISDN

30  channels.

WO 97/37501 describes a telecommunications systems with agent-negotiated routing.

More generally, however, the extension of the principles of the present invention to resource allocation systems in which individual items of work are distributed (rather than rates at which work is to be done) is not excluded.

A particularly convenient format for providing the software modules is in

5   the form of software agents.

This can be provided as an arrangement referred to herein as "Agent Enhanced Workflow" (AEW). Agent Enhanced Workflow is a novel technique whereby a community of intelligent, distributed, autonomous software agents is used to improve the management of business processes currently under the

10   control of Workflow Management Systems. Agent Enhanced Workflow can specifically address shortcomings in such systems by adding a layer of software agents to the Workflow Management System architecture. These Workflow Agents can collaborate to perform real-time exception handling and/or co-ordinate the distribution of work items. This is to be distinguished from "Agent Based

15   Workflow", which replaces such existing architecture as in the prior art cited above.

The layer of software agents can be added retrospectively to an existing workflow management system or can be designed into a new workflow management system as an integral part.

20   Software agent technology has developed over the past few years in several different fields. A software agent is a computer program which acts as an agent for an entity such as a user, a piece of equipment or a business. The software agent usually holds data in relation to the entity it represents, has a set of constraints or conditions to determine its behaviour and, most importantly, is

25   provided with decision making software for making decisions on behalf of the entity within or as a result of the constraints and conditions. Agents are generally acting within a system and the decisions an agent makes can result in activity by the system. In control software systems, those decisions result in control activity, such as initiating connection set-up in a communications network

30   controlled by the system.

An agent acting within a system will also generally hold data about the system so that it can operate in context.

4

In a distributed environment, many such agents may co-operate to co-ordinate and perform the control activities. Typically, such agents form an agent layer, with each agent interfacing with one or more external systems (the domain layer) which they control, monitor or manage.

5    An agent-based system can be very complex since the interactions between the agents, the decision-making processes of individual agents and the interactions between agents and the external systems they control, need to be taken into account.

Different types of agent-based systems are described in several published
10  papers, such as those published in the proceedings of the First and Second International Conferences on the Practical Application of Intelligent Agents and Multi-Agent Technology. These are published by the Practical Application Company Ltd., Blackpool, Lancashire, in 1996 and 1997 respectively. A general comprehensive review of agent-based technology is given by Hyacinth S. Nwana,
15  "Software Agents: An Overview" in the *Knowledge Engineering Review* journal, Vol. 11, No. 3, pages 205-244.

Returning to AEW, an agent layer in a workflow management system can in particular overcome the inability of a workflow management system (WfMS) to cope with dynamic changes in resource levels and task availability. Similarly the
20  Workflow Agents can co-ordinate overall resource levels, bringing individual resources on and off-line as required to accommodate peaks and troughs in the incoming workload.

A workflow management system for a correspondence handling centre (CHC), according to an embodiment of the present invention, will now be
25  described, by way of example only, with reference to the accompanying Figures in which:

Figure 1 shows the overall organisation of components of the correspondence handling centre, together with its interfaces to a user community;

Figure 2 shows an overview of a process definition for the CHC;
30  Figure 3 shows the relationship between logical components of the workflow management system for the CHC;

Figure 4 shows a schematic view of the architecture of a workflow management system agent;

Figure 5 shows schematically the technological components of the CHC and the agent-based workflow management system;

Figure 6 shows schematically a protocol for negotiation between the agents of the workflow management system;

Figure 7 shows a price profile for use in the collaborative processes of the CHC agents;

Figures 8, 9 and 10 show steady state and negotiating phases of interactions between agents of the CHC;

Figure 11 is a block diagram showing the arrangement of computers on which the embodiment is implemented;

Figure 12 (comprising Figures 12a-12d) is a flow diagram showing schematically the operation of a central administration agent; and

Figure 13 is a flow diagram showing schematically the process of operation of a work processing centre agent of the embodiment.

### FIRST EMBODIMENT

### GLOSSARY

AEW: Agent Enhanced Workflow

CA : Central Administration

CAA : Central Administration Agent

CBR : Case Based Reasoning

CHC : Correspondence Handling Centre

SLA : Service Level Agreement

WAPI : Workflow APIs and Interchange Formats

WfMS or WMS : Workflow Management System

WPC : Work Processing Centre

WPCA : Work Processing Centre Agent

WMC : Work Management Centre (CA plus one or more WPCs)

WMCA : CA Agent(s) plus WPC Agents

## DEFINITIONS

For the following and other definitions see 'Terminology & Glossary', Workflow
Management Coalition (MFMC-TC01011), June 1996, 2.0.

5

Broadcast (message) : sent to all WPC Agents associated with a given CA Agent

Cost: a measure of the difficulty of achieving a particular objective

Delta: a vector quantity representing some change to the parameters of one or
more of the work item categories processed by the Workflow

10    Management System

Narrowcast (message): sent to a subset of the WPC Agents associated with a
given CA Agent

Normal capacity of WP : the maximum rate of work which a WPC can achieve,
without recourse to additional, short-term staffing arrangements

15  N-way routing: synonymous with 'Parallel Routing', 'Parallel workflow processing'
and 'Concurrent Processing'. "A segment of a process instance under
enactment by a workflow management system, where two or more
activity instances are executing in parallel within the workflow, giving rise
to multiple threads of control"

20  Price: the monetary value (in pounds and pence or some other suitable unit)
offered/charged for delivery of a specified service

Rate of Work: the volume of work per unit time

Service Level Agreement: an agreement between a WPC and the CA, which
commits the WPC to undertaking a given work profile and the CA to

25    supplying it. Price, rate, quality etc. parameters are specified for each
work item category present in the SLA. Note that SLAs refer to an on-
going agreement, rather than a one-off

Steady state operation : the WMC is in equilibrium (the rate at which work leaves
the WMC is equal to the rate at which work enters the WMC)

30  Volume of Work : amount of work

Process definition : "The representation of a business process in a form which
supports automated manipulation, such as modelling, or enactment by a
workflow management system...."

Workflow: "The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules."

Workflow Management System: "A system that defines, creates and manages the
5 execution of workflows through the use of software, running on one or more workflow engines, which is able to interpret the process definition, interact with workflow participants and, where required, invoke the use of Information Technology (IT) tools and applications"

Workflow Enactment Service: "A software service that may consist of one or
10 more workflow engines in order to create, manage and execute workflow instances. Applications may interface to this service via the workflow application programming interface (part of WAPI)."


## CORRESPONDENCE HANDLING CENTRE

15

Referring to Figure 1, the correspondence handling centre (CHC) 120 which provides the domain for the workflow management system described below is that of an enterprise receiving a stream of correspondence from its customers concerning its service offerings. These requests are fed into the
20 enterprise's Correspondence Handling Centre (CHC), an entity which may incorporate not only in-house processing facilities but also (or instead) one or more processing facilities that are operated by third-parties (i.e. out-sourced).

Referring to Figure 1, the Correspondence Handling Centre 120 is composed of a number of disparate Work Processing Centres (WPCs) 100 and a
25 Central Administration (CA) 105. The CA 105 provides an interface to a user community for the receipt of work requests 115 and for issuing reports 125.

The CA 105 deals with all stages of the CHC business process, bar the processing of the individual work items, which is performed by the WPCs 100. An individual WPC 100 may either be co-located with the CA 105, or sited at a
30 remote location. In practice a WPC 100 may act for more than one CA 105 but in the scenario discussed here, a WPC 100 will be associated with only one CA 105, and hence a single CHC. Similarly a CHC may service more than one

enterprise or an enterprise may use more than one CHC but again, a one-to-one relationship is represented here.

A CHC may handle all or part of the correspondence received by the enterprise it serves. This correspondence can be of many types, ranging from
5  requests to quote for new business, through complaints about existing goods or services, to requests to modify or remove/cease goods or services already provided.

**CHC Processes**

10

Referring to Figure 2, the processes carried out by the CHC can be simply represented in terms of six specific categories of activity, each of which is described briefly here (note that, of the six, only 'Processing' is performed by the WPCs 100, the rest being the concern of the CA 105):

15

1. Reception 230:
Correspondence arrives at the CA 105 either in batches in the post or asynchronously by telephone, electronic mail, facsimile and/or similar means. Work arrival patterns and content are determined by factors outside of the control
20  of the CHC.

2. Classification 225:
Work is typically classified by type and difficulty. A given piece of work may contain items of more than one type, requiring it to be decomposed into a number
25  of separate work items for classification, distribution and processing. These work items are then recombined before being passed on to Inspection.

3. Distribution 205:
Work items are distributed to the WPCs 100 for processing. This distribution
30  activity is based on a complex mix of parameters which are used to match work items with available, appropriately skilled resources at the WPCs 100.

4. Processing 210:

Work items are processed at the WPCs 100, and are either returned for checking (if completed), or returned unprocessed for redistribution. Each WPC 100 has variable resource levels - it may be able to call in part-time staff at short notice or

5    put the regular staff onto overtime as and when required. Each WPC 100 also has its own operational profile due to the attributes of the resources available to it, and its overall business objectives.

5. Inspection 215:

10   Completed items are checked and reworked if necessary. Typically, the inspection activity will have one of three possible outcomes:

* major fix required - return item to WPC 100 for re-work;

* minor fix required - typically, fix item locally and forward to Dispatch;

* no fix required - forward item to Dispatch.

15

6. Dispatch 220:

Completed items are dispatched to the customer and records may be kept for future billing, customer query and auditing purposes.

20         Workflow management systems are known which will deal with this type of domain but they also have inherent problems. In particular:

* the inability to cope with dynamic changes in resource levels and task availability;

25   * a lack of resource management facilities;

* inadequate exception handling, especially during the processing of decomposed items;

* a limited ability to predict changes, due to external events, in both the type and volume of work entering the business process;

30   * the inability to improve dynamically both the business process and how it is managed;

* a limited or non-existent ability to manage the decomposition and recombination of complex items.

These shortcomings manifest themselves as a mismatch between the
actual capacity of a WPC 100 and the work offered to it at a given time. This in
turn leads to sub-optimum throughput, necessitating wasteful over-provisioning of
resources and/or backlogs of work. The present embodiment reduces some of
5    these problems

## SYSTEM ARCHITECTURE AND TECHNOLOGY

10    Referring to Figure 3, an embodiment of the present invention comprises
an agent layer 300 which forms a layer above a Workflow Management System
305 which in turn controls the actual processing components 100 in the
processing domain 310 of the correspondence handling centre. The CHC 120
consists of the WfMS 305 and a collection of tasks and resources, including the
15    WPCs 100, in its processing domain 310.

The agent layer 300 comprises at least one central administrative agent
320, and several work processing centre agents 325. Together the agents can
determine resource allocation for dealing with workflow and in particular can deal
with two different types of situations:
20

* External exception - This occurs when there is a change in the overall workload
presented to the CHC 120. The CA Agent 320 is informed and initiates a dialogue
with the WPC Agents 325.


25    * Internal exception - This situation arises when a resource exception occurs at a
WPC 100 (for instance a 'flu' epidemic may render a large part of the workforce
unfit for work). This exception is reported to the appropriate WPC Agent 325 by
the WPC 100 in the processing domain 310, and the WPC Agent 325 instigates a
negotiation with the CA Agent 320.
30

In both cases, the agents of the agent layer 300 attempt to form a
distribution plan, to show how required changes to overall workload or current
work distribution could be achieved.

The CAA's chief function is to co-ordinate the distribution of work across the processing centres (the WPCs 100) in the most efficient manner possible. This is achieved via negotiations between the CAA 320 and the various WCPAs 325 - the aim being to establish new contracts to cover the changed workload.

5 Note that there is no inter-WPCA communication at any stage, as the WPCAs 325 are only aware of the existence of the CAA 320, and not the other WPCAs that form part of a given AEW environment.

All the Workflow Agents 320, 325 are autonomous so they function independently when negotiating for new contracts, or for changes to existing ones. Each WPCA 325 has knowledge of the work commitments of its own WPC 100, and negotiates to obtain a workload that will ensure its resources are utilised in the most efficient manner possible. Both the CAA and the WPCAs are attempting to meet their individual business objectives, and the details of the negotiations are geared towards achieving these ends.

15

AEW Technology Overview

Referring to Figure 5, the WfMS implements the CHC business process; the CHC domain drives work through the WfMS in accordance with the current
20 version of the distribution plan.

A three-tier architecture can be seen, with a Frontware layer of information and control GUIs, a Middleware layer comprising the distribution mechanism, and the Backware layer made up of the functional engines of the various components.

25 Backware: In order to allow platform independence, the Workflow Agents are implemented in Java using the JDK1.1 development kit. The WfMS is a commercially available third-party product.

Middleware: This is based upon a commercially available, CORBA 2-compliant, distributed computing platform, and comes in two different flavours:
30 Java to Java (for inter- and intra-agent, and agent to Frontware communication); Java to C++ (for agent to WfMS and agent to CHC domain simulator communication).

Frontware: The Workflow Agents use GUIs implemented in Java to provide administration and reporting tools to the end user. The WfMS uses proprietary administration tools, running on a PC platform. Multi-threading allows a single client to make concurrent calls to the WfMS, and the same call to be
5    made by a number of clients concurrently. This means that all the Workflow Agents are able to monitor work as it passes through the WfMS and the CAA can update the contents of the SLAs as and when its negotiations with the various WPCAs are concluded.

Referring to Figure 11, the software architecture of Figure 5, is
10   implemented in this embodiment by four computers 1000; 1100; 1200; 1300. The CAA agent is provided on a first computer 1000. The location of this computer may, for example be close to that of the customer supplying work, or the headquarters of the call handling centre.

The computers 1100, 1200, 1300 are conveniently, as shown, the same
15   computers on which the work processes 305 are performed. Thus, the work process agents 100 are co-located with the work to be performed, and are arranged to receive messages therefrom, in the format defined by the workflow management system to which the work processes confirm (i.e. the above mentioned standard) indicating the changes in the ability of the work process to
20   do work. Conveniently, the computers 1100, 1200, 1300 comprise keyboards (not shown) via which human operators of the computers, performing the correspondence answering process, can input data indicating a change in their ability to work (for example their presence or absence).

25   **Workflow Agent Architecture**

Referring to Figure 4, both types of Workflow Agent currently implemented have the same basic architecture. However, their business objectives and the details of their constituent modules differ.
30   Both the CAA and WPCA communication modules 400 perform similar functions - they enable their respective agent to communicate with the outside world (i.e. with other Workflow Agents, the underlying WfMS and humans, via GUIs), and provide a channel for intra-agent communication.

The Workflow Agents' Environment Models 405 contain knowledge of their own operational characteristics; the Collaboration Modules 410 contain the algorithms and strategies that are used while negotiating for contracts; and the Co-ordination Modules 415 are responsible for all aspects of the Workflow
5    Agents' functionality other than communication and collaboration.

Service Level Agreements (SLAs) are stored in the environment model 405, together with residue queues. These are further discussed below.

**Service Level Agreements**

10

The purpose of an SLA is to record an agreement between the CA 105 and an individual WPC 100. The agreement is binding in both directions, in that the CA undertakes to provide a certain work profile at a given rate, and the WPC undertakes to process it. All of the agents make use of a common information
15    model and understand the same thing about work categories, rates etc.

Thus the notion of service is entity dependant: a WPC 100 receives a work-item provision service from the CA 105; the CA 105 receives work-processing services from its WPCs 100. These services are on-going, in that each has a duration (which may be infinite - which in effect means 'until further notice') and a start
20    time (which allows a service to be planned in for commencement at a later date/time).

An SLA includes the parameters detailed in Table 1. Given this format for an SLA, it follows that there will be a number of SLAs in place between each WPC 100 and the CA 105 (i.e. one for each work category).

25

| Identity | The SLA's unique ID |
|---|---|
| CA Id | Unique identifier of the CA bound by this agreement |
| WPC Id | Unique identifier of the WPC bound by this agreement |
| Work category | An entry indicating the type of work for which this SLA is in place |
| Rate | The volume of work items per time unit (however defined) that the WPC Agent has agreed to process |
| Quality | Quality control parameter (e.g. % of jobs that will require no rework) |

14

| Price | The price (paid by the CA) for these work items to be processed |
| Recall Penalty | The price paid by the WPC Agent, if it is unable to process the work items specified in this SLA |
| Start Time | The time at which the service will be available |
| Duration | The length of time the service will be provided for |

**Table 1: The structure of an SLA**

SLAs are updated accordingly to reflect any negotiated changes to the
5   distribution of work amongst the WPCs 100.

**CAA-Specific Architecture**

The CAA's Environment Model 405 maintains records relating to all work
10  submitted to the CHC for processing. These records cover three separate
categories: allocated, unallocated and unavailable batches of work. The records
for unallocated and unavailable batches can be in the form of "residue queues".

Allocated batches - the Environment Model 405 is aware of all work that
has been allocated as a result of successful inter-agent negotiations, as it holds a
15  local copy of all the SLAs that have been placed with the WPCAs 325.

Unallocated batches - It also records work, in the form of a list or queue,
that has been submitted to the CHC but could not be distributed at the first
attempt because insufficient processing capacity was available at that time. The
WPCAs 325 are at liberty to under-bid (bid for less work that the bid request
20  specified) during any phase of the negotiation, which may lead to a cumulative
under-bid across the CHC as whole. The CAA 320 has no choice but to record
any such under-bid, and attempts to distribute the resulting residual work by
combining details of its type and rate with the contents of the next delta received.
This is done with a view to forming an enlarged bid request for dissemination to
25  the WPCAs 325 in the usual fashion.

Unavailable batches - Similarly, the CAA's Environment Model keeps track
of internal requests for work that could not be satisfied at the time. These
requests are communicated via over-bids (bidding for more work than the bid

request specified) from individual WPCAs to the CAA. The CAA attempts to satisfy these requests by checking the records relating to unallocated batches and adding any matching (in terms of rate, quality and processing period, by work type) items to the distribution plan, then updating the list of unallocated batches.

5   If the request cannot be satisfied in this fashion, the CAA notes the request, and attempts to satisfy it upon receipt of the next delta. If the delta contains items that match the request, these are added to the next distribution plan, and omitted from the subsequent bid request sent out by the CAA.

The unallocated and unavailable batches mentioned above can be stored

10  by the CAA as "residue queues" which can be "excess" or "deficit" queues.

The other key function of this Environment Model is to provide both directory information and name server functionality so that the CAA can locate and communicate with the WPCAs that act for the WPCs belonging to the CHC.

The Co-ordination Module determines the algorithm used to re-formulate

15  the distribution plan, in an attempt to accommodate successive changes in the amount of work that can be processed by the CHC as a whole. It has recourse to the details of any unallocated or unavailable batches, and implements the attempts to match these to portions of the most recently received delta.

The Collaboration Module executes the collaboration strategy used by the

20  CAA to obtain offers from the WPCAs for changing the contract governing the workload of their WPCs.


**WPCA-Specific Architecture**


25      The WPCA Environment Model 405 is responsible for maintaining records of work it has agreed to accept from the CAA 320. It does this by holding local copies of any current contracts (SLAs) it has successfully negotiated for. It is also responsible for maintaining its resource model and price profile.

Its Co-ordination module 415 gathers together all information relating to

30  the WPC's ability to process work - this must be up to date in order to allow the WPCA 325 to make realistic, timely responses on receipt of bid requests from the CAA 320.

The Collaboration Module executes the collaboration strategy used by the WPCA 325 to generate responses to the CAA's requests to change the contracts governing the workload of its WPC 100.

## 5 Deltas

The Workflow Management System will normally be in a steady state. Any steady state is likely to be temporary however, as the requirement for work may increase or decrease over time, and resource levels at the WPCs 100 can 10 change. Such a perturbation is referred to as a delta (or '$\Delta$').

The delta is a vector quantity, and includes the parameters show in Table 2.

| $\Delta$identity | | | | | |
|---|---|---|---|---|---|
| Work category | Source | Rate | Quality | Start Time | Duration |
| Work_Cat_1 | External | $\Delta v_1$ | $\Delta q_1$ | $st_1$ | $d_1$ |
| Work_Cat_2 | External | $\Delta v_2$ | $\Delta q_2$ | $st_2$ | $d_2$ |
| . | . | . | . | . | . |
| Work_Cat_N | External | $\Delta v_N$ | $\Delta q_N$ | $st_N$ | $d_N$ |

15 Table 2: Details of a Delta

An entry of the form $\Delta v_i$ indicates an increase/decrease in the rate of work category I. Similarly $\Delta q_i$ indicates a change in the quality of work I. Note that $\Delta v_i$ or ($\Delta q_i$ may be zero (i.e. no change to rate/quality of that type of work 20 item). There will be an entry for each category of work handled by the system, even if no change is required. The source of a delta can be either 'External' (to indicate that it originated from outside of the WMC) or the identifier of a particular WPC (to identify the WPC that is requesting a change to its current workload). An example external $\Delta$ is shown in Table 3, and an example internal $\Delta$ is shown in 25 Table 4.

Deltas can be 'positive' and 'negative'. A positive delta is one in which all the non-zero rate values are positive (e.g. Table 3), whereas a negative delta is

one in which all the non-zero rate values are negative (e.g. Table 4). A mixed
delta is one which contains a mixture of both positive and negative non-zero rate
values. It follows that deltas may be broadly characterised as being either external
or internal, and either positive or negative or mixed.

5

| ΔIdentity = Δ1 Work category | Source | Rate | Quality | Start Time | Duration |
|---|---|---|---|---|---|
| A | External | +30 | 0 | NOW | INFINITE |
| B | External | 0 | 0 | NOW | INFINITE |
| C | External | +20 | +10% | NOW | INFINITE |

Table 3: An example external delta

| ΔIdentity = Δ2 Work category | Source | Rate | Quality | Start Time | Duration |
|---|---|---|---|---|---|
| A | WPC2 | -10 | 0 | NOW | INFINITE |
| B | WPC2 | 0 | 0 | NOW | INFINITE |
| C | WPC2 | -5 | 0 | NOW | INFINITE |

10   Table 4: An example internal delta


OVERVIEW OF WORKFLOW AGENT COLLABORATION


        Referring to Figure 6, the co-ordination strategy used by the Workflow
15   Agents in this scenario uses a combination of Contract Net and limited Contract
Net protocols. These are published in: 'Negotiation as a metaphor for distributed
problem solving' by Davis R and Smith RG, in Artificial Intelligence, No 20, pp63-
109 (1983), and: 'The Contract Protocol: High-Level Communication and Control
in a Distributed Problem Solver', by Smith RG, IEEE Trans. on Computing, 29, No
20   12 (December 1980). The difference between the two is that Contract Net
always announces the start of the bidding process to all available contractors,
whereas limited Contract Net announces to a specific sub-set of contractors. The
Workflow Agents 320, 325 combine both by opening the negotiation process by

18

inviting all associated WPCAs to bid, then progressively excluding successful bidders, as shown in Figure 6.

In Contract Net terms, the CAA 320 assumes the role of the human manager by dividing the problem (work distribution) into sub-problems
5  (distributing work items), searches for contractors to carry out tasks and monitors the overall solution by maintaining local copies of all its contracts. The WPCAs 325 assume the role of human representatives of the contractors who carry out sub-tasks (processing work items). This is not a 'zero-sum game', that is the WPCA 325 can increase the profits of its WPC 100 at the same time as helping
10  the CAA 320 to increase the profits of the CA 105. Agents in the role of managers locate suitable contractors via an activity of bidding which proceeds as follows (taken from: 'Co-ordination in Multi-Agent Systems', by Nwana HS, Lee L and Jennings NR, in Nwana H S and Azarmi N (Ed): 'Software Agents and Soft Computing - Towards Enhancing Machine Intelligence', Springer, pp42-58
15  (1997)):

* a manager announces a task;
* contractors evaluate the task with respect to their own abilities and commitments;
20  * contractors table bids to the manager;
* the manager evaluates received bids, chooses a contractor and awards the contract to it;
* the manager waits for the result of the contract.

25      The following section contains simple, worked examples for both an external and an internal delta. Only work rate and price attributes are shown, and not all of the possible search space is explored. It is assumed that no over-bids are received in either case.

There are three WPCAs 325 used in these examples (WPC1A, WPC2A
30  and WPC3A) and three work item types designated A, B and C.

### Handling an External Delta (Including Example)

Upon receipt of an external delta, the CAA 320 multi-casts (i.e. sends a message addressed to all appropriate recipients) a bid request to the WPCAs 325.
5  This bid request contains various details for each category of work that the CAA 320 is trying to distribute, including required start and end times, work rate (i.e. items per unit time) and quality threshold. Each WPC Agent 325 then uses its Co-ordination module to determine its response in terms of quantity, quality, price etc. for each offered work category. The WPCA's Collaboration Module 410
10  sends this bid response to the CAA 320, though it can decline to respond if it cannot process any additional work in the time period specified in the original bid request.

As well as bidding for some or all of the offered work, the WPCA 325 may think it prudent to overbid, i.e. bid for more work than was actually offered
15  by the CAA 320. Overbidding might occur when a WPCA cannot fulfil a bid request unless it brings additional resources on-line. In order to fully utilise these extra resources, it may require more work than is currently on offer, hence it overbids. If accepted, the CAA records the details of the overbid, with a view to offsetting it against the contents of subsequent deltas.

20  The CAA 320 either waits for replies from all of the WPCAs 325 to arrive or for some pre-determined time-out period to expire. Then the cost factor (which takes into account the bid price, difference in bid and response quantity, quality etc. for each work category included in the initial bid request) is calculated for each bid received, and the bids are ranked by ascending cost.

25  The lowest 'cost' bid is used as the basis for the next round of negotiation (see below for a discussion of 'cost' and 'cost factors'). If this bid accounts for all the offered work, then this branch of the negotiation is terminated. Otherwise the CAA 320 narrowcasts a new bid request message to cover the difference between the quantity (per unit time) of each work type
30  covered by the bid response with the lowest cost factor, and the requested rate for each work category contained in the original, multicast, bid request. This difference represents the work that would not be distributed if the CAA 320 simply accepted a single, lowest cost bid. The WPCA 325 that supplied the

lowest cost bid is excluded from the list of recipients. The CAA 320 receives another set of bid responses, ranks them in ascending cost as before, and calculates the difference between requested and offered work rates, also as before. This process continues until either all of the work has been accounted for,

5    or there are no more WPCAs 325 to solicit bids from. Any outstanding work is recorded for distribution at a later time.

The CAA 320 has now established a baseline cost for placing some or all of the work contained in the external delta. It now repeats the 'multicast, followed by exhaustive narrowcast' process for each of the remaining, initial bid

10   responses. However, it can now prune the search space by comparing the cumulative cost, after each round of bidding, with the baseline cost - the current branch of the search is abandoned if the baseline cost is exceeded.

Once the lowest cost, available (as opposed to ideal) solution has been found, the CAA 320 records the details of the new contracts that it has

15   negotiated and forwards the updated distribution plan to the WfMS 305 for implementation.

Suppose the following (simplified) external delta arrives at the CA Agent:
(20A, 10B, 5C)

This indicates that an additional 20 work items of type A per unit time,

20   10 of type B and 5 of type C are arriving at Reception.


Contract Net Phase

A multicast bid request covering this additional work is sent out to all
25   WPC Agents:

"Can you process (20A, 10B, 5C) extra work items per unit time?"


The WPC Agents return the following offers:
"(x1A, y1B, z1C) @ price U"      from WPC1A;
30   "(x2A, y2B, z2C) @ price V"      from WPC2A;
"(x3A, y3B, z3C) @ price W"      from WPC3A.


This indicates to the CAA 320 that:

WPC1A can accommodate an additional x1 work items of type A per unit time, y1 work items of type B and z1 work items of type C at a total price of U,

WPC2A can accommodate an additional x2 work items of type A per unit time, y2 work items of type B and z2 work items of type C at a total price of V, and

5    WPC3A can accommodate an additional x3 work items of type A per unit time, y3 work items of type B and z3 work items of type C at a total price of W.

The cost factor is calculated using a comparative function, the general form of which is:

$$CF = w1f1(x) + w2f2(x) \ldots + wnfn(x)$$

10    where each $fn(x)$ represents the evaluation of a quantitative aspect of the bid response (i.e. rate, quality, timeliness), and each $wn$ represents the relative importance of a term. In practice it was useful to enforce the following constraint:

$$w1 + w2 + \ldots + wn = 1$$

The weights are used to reflect the current business priorities of the CAA

15    320. In this example we enforce a 'maximum work distribution is much more important than price, all other factors are ignored' policy, by setting

the rate weight $w1 = 0.8$,

the price weight $w2 = 0.2$ and

all the other weights to zero.

20    The cost factor for each bid response is now calculated by the CAA:

$$CF1 = 0.8( (20\text{-}x1) + (10\text{-}y1) + (5\text{-}z1) ) + 0.2U \qquad \text{for WPC1A's initial bid}$$
$$CF2 = 0.8( (20\text{-}x2) + (10\text{-}y2) + (5\text{-}z2) ) + 0.2V \qquad \text{for WPC2A's initial bid}$$
$$CF3 = 0.8( (20\text{-}x3) + (10\text{-}y3) + (5\text{-}z3) ) + 0.2W \qquad \text{for WPC3A's initial bid}$$

25    If WPC1A's bid has the lowest cost factor, and WPC3A's the highest, and none of the responses exactly matched the work rates contained in the initial bid request, then the negotiation proceeds, starting from the lowest cost (i.e. WPC1A's) bid response.

30  **Limited Contract Net Phase**

A narrowcast bid request is sent out to WPC2A and WPC3A:

"Can you process ( (20 - x1)A + (10 - y1)B + (5-z1)C ) extra work items per unit time?"

The bid responses are as follows:

"(x21A, y21B, z21C) @ price X"   from WPC2A;

5    "(x31A, y31B, z31C) @ price Y"   from WPC3A.

The cost factors are calculated as before:

CF21 = 0.8( (20-x1-x21 ) + (10-y1-y21) + (5-z1-z21) ) +0.2X for WPC2A's second bid

CF31 = 0.8( (20-x1-x31) + (10-y1-y31) + (5-z1-z31) ) +0.2Y for WPC3A's

10   second bid

Provided that neither bid response exactly matched the work rates contained in the second bid request, and assuming that CF21 < CF31 (i.e. WPC2A's bid was the 'best' of the two), a final narrowcast message is sent (in this case to WPC3A only):

15   "Can you process ( (20-x-x21 )A + (10-y-y21)B + (5-z-z21)C ) work items per unit time?"

The final bid response of this round of the negotiation is received:

"(x321A, y321B, z321C) @ price Z" from WPC3A.

The baseline cost is calculated by summing the cost of the individual

20   successful bids:

Costbaseline = 0.8( (20-x-x21-x321 ) + (10-y-y21-y321) + (5-z-z21-z321) ) + 0.2(U + X + Z)

(Note that the intermediate branches are also explored in ascending cost factor

25   order, but that this has been omitted from the example for the sake of brevity).

Subsequent Iterations

The Limited Contract Net phase is repeated twice, commencing each time

30   with the next 'best' bid received from the initial Contract Net phase. The difference being that after each step, the cumulative cost is compared with the baseline cost, and the negotiation is abandoned if the baseline cost has been

23

exceeded. Once the CAA has determined the lowest cost solution, it places contracts with the successful bidders and records the work rate residue (if any).

### Handling an Internal Delta

5

When a WPCA 325 is alerted that its WPC's 100 ability to process work has been reduced for any reason (e.g. 'flu' epidemic, local public holiday) it generates an internal delta. This contains details of the types and rates of work that the underlying WPC 100 wishes to shed (in WfMS terms, return unprocessed

10 to the CA 105), along with the details of the period for which shedding is requested, and is sent to the CAA 320. Upon receipt, it is converted to the equivalent of an external delta, by the simple expedient of reversing the signs of the work rate values. The negotiation protocol is then very similar to that for an external delta, the one difference being that the initial bid request message is not

15 sent to all WPC Agents as before, rather a narrowcast is sent to all WPC Agents bar the one that raised the internal delta.

Suppose WPC2A generates a request to reduce its type B workload by 5 items per unit time. It sends the following internal delta to the CAA (all other details omitted):

20          (0A, -5B, 0C)

### Limited Contract Net Phase

The CA Agent reverses the polarity of any non-zero rates, checks its

25 records for work owed due to previous over-bidding and, if the delta cannot be offset from this, sends a narrowcast message to the other two WPCAs:

"Can you process (0A, 5B, 0C) extra work items per unit time?"

The bid responses are as follows:

30  "(0, y1B, 0) @ price X"          from WPC1A;

"(0, y3B, 0) @ price Y"          from WPC3A.

The cost factors are calculated as before:

24

$$CF1 = 0.8( (5-y1) ) + 0.2X \qquad \text{for WPC1A's initial bid}$$
$$CF3 = 0.8( (5-y3) ) + 0.2Y \qquad \text{for WPC3A's initial bid}$$

Provided that neither bid response exactly matched the work rates contained in the initial bid request, and assuming that CF1 < CF3 (i.e. WPC1A's bid was the 'best' of the two), a final narrowcast message is sent (in this case to WPC3A only):

"Can you process ( (5-y1)B ) work items per unit time?"

The final bid response of this round of the negotiation is received:

"(y31B) @ price Z"         from WPC3A.

The baseline cost is calculated by summing the cost of the individual successful bids:

$$Costbaseline = 0.8( (5-y1-y31) ) + 0.2(X + Z)$$

## Subsequent Iterations

The Limited Contract Net phase is repeated, but using WPC3A's initial bid this time. Again, the cumulative cost is compared with the baseline cost, and the negotiation is abandoned if the baseline cost is exceeded.

Once the CAA 320 has determined the lowest cost solution, it places contracts with the successful bidders and records the work rate residue (if any). It then updates WPC2A's contract to record the fact that the type B work rate has been reduced by 5 items, for the period specified (but not shown here) in the internal delta.

A description of the algorithm will now be given in greater detail with reference to Figures 12a to 12d.

In step 2002, as described above, the CA agent broadcasts the original amount of work to all WPCAs.

In step 2004, the process of Figure 12b is called. Referring to Figure 12b, in step 2006, replies are received from the WPCAs. In step 2008, the routine of Figure 12c is called.

Referring to Figure 12c, in step 2010, for each of the replies, a running price total is calculated comprising the sum of the price specified in the reply, and

that previously calculated in the last stage of the algorithm (which is zero for the first round).

In step 2012, the running price is compared with the stored baseline price (if any), and if the price is greater than the baseline price then that branch is
5   rejected and not further considered (step 2014).

Next, in 2016, it is determined whether each of the returned quantities matches exactly the transmitted quantity (step 2016) and, if so, the price (which, it will be recalled, is lower than the existing baseline price) is made the new baseline price and the details of the corresponding contracts with the WPCAs
10  making up the running price are stored as the baseline allocation (step 2018).

If the quantity concerned did not exactly match (i.e. was less than) the quantity transmitted, then as described above the cost factor for the bid is calculated (step 2020).

If the last received signal has been processed (step 2022), the process of
15  Figure 12c returns; otherwise (step 2024) the next received bid signal is processed.

Thus, the process of Figure 12c exits with the lowest received exact allocation stored as the baseline allocation and its total price stored as the baseline price; and with all partial solutions having associated cost factors
20  calculated.

Returning now to Figure 12b, in step 2030, any such partial bids are ranked by cost factor. On having reached the last partial bid (i.e. if there are none left to consider) the process of Figure 12b exits. Figure 12b will therefore exit at the end of a given round of signalling (step 2032).

25         If not, then the lowest cost remaining partial allocation to be considered is selected (step 2034) and the routine of Figure 12c is selected.

Referring to Figure 12c, in step 2040, the remaining source quantity which would be required if the previous partial bid were used as a basis is calculated, by adding to a running total of the quantities previously bid in
30  preceding rounds the amount bid in the bid being considered and subtracting this from the total amount original required (step 2040).

In step 2042, this remaining amount is transmitted to all WPCAs other than that which gave rise to the bid now being considered. (For the avoidance of

doubt, it will be understood that these include WPCAs which have previously made bids on previous rounds. All such previous bids from WPCAs are, at this stage, neither accepted or rejected but "on hold" while the CAA evaluates the best allocation).

5          Next, in step 2044, the routine of Figure 12b is called once more, to evaluate this next round of bids. It will be understood that, by using recursion, each of the routines of 12b, 12c, and 12d can be called from within itself, partial results being stored on stacks until the return from each level of recursion.

          On completing execution of the process of Figure 12b, the process of
10  Figure 12c returns.

          On the final return from the process of Figure 12b, referring to Figure 12a, in step 2050 the CAA determines whether (as should be the case) a baseline allocation (comprising one or several (price, quantity) offers from WPCAs) is stored. If so, in step 2052, the CAA signals, to each WPCA, acceptance of that
15  (price quantity) offer (step 2054) and the details are recorded as described above.

          If the amounts of the allocation do not match the original requirement, or if no baseline was derived, then the residue is stacked on a residue queue (step 2056) and added into the next received external delta.

          Thus, the effect of the algorithm is to perform a depth-first search for
20  solutions, taking each of the replies at each stage of negotiations for each branch, in turn.

          Whereas this would potentially result in a very large number of communications between the agents, requiring substantial bandwidth and/or long delay in determining a resource allocation, by the use of a baseline price (which is
25  preferably updated as successively better solutions are found) more and more possible branches are excluded as the search proceeds.

          It will be understood that substantially the same algorithm is performed when an internal delta is received, except that the work processing centre agent which gave rise to the internal delta is not contacted during at least the first
30  round (and possibly further rounds) of the negotiation process.

          The above described process is preferably improved, as discussed above, by further modifications to reduce the number of solutions evaluated (for example, the signalling of a maximum price by the CA agent as discussed above).

WMC AGENT OPERATION

The WMC Agent's aim is to enhance the distribution process by
5   establishing formal agreements between the WPCs and the CA (i.e. the SLAs) that
are intended to regulate the flow of work items through the Workflow System.
These negotiated agreements underpin each distribution plan that the CA Agent
proposes to the User. Any proposed changes to an existing SLA requires the
consent of both the CA 105 and the WPC 100 in the first instance, and eventual
10  approval by the User.

The CA Agent 320 could form a centralised distribution plan in isolation
(and consequently the WPC Agents 325 would have a dubious claim to
agenthood, due to their lack of autonomy), if it had detailed knowledge of the
resources, workload, price profile etc. of each WPC 100. However it is more
15  realistic to assume that this knowledge is preferred not be widely available.
Therefore direct access to it is restricted to the agent(s) responsible for a given
WPC 100.

In order for the negotiation to proceed, the WPC Agents are asked
questions such as the following;
20  1. "Can you increase your rate of type A work by 30 items per unit time from
<start time> for <duration>?"
2. "Can you increase your rate of type A work by 30 items per unit time from
<start time> for <duration > for a price less than or equal to 50?"

For a WPC Agent to be able to answer these questions, it must have
25  access to knowledge about the capabilities of its WPC, such as the rate and
profile of work that the WPC is committed to. It is assumed that it will have
access to the details of the available workforce, such as the number of workers,
their skills, rates of pay, and other factors that influence the price profile.

All of these factors will be brought to bear in a complex calculation which
30  will result in an answer of one of the following forms:
1. "I can increase my rate of type A work by 30 items from <start time> for
<duration> at a total price of 34"

2. "I can increase my rate of type A work by 25 items from <start time> for <duration>, and the price for this will be 20"

3. " I can increase my rate of type A work by 40 items from <start time> for <duration> at a total price of 25"

5    Each response can be categorised as either an underbid, an matching bid or an overbid. Note however that 'under', 'match' and 'over' refer to individual attributes of the initial query (e.g. rate or quality) but not all (i.e. rate, quality, start time, duration or price).

* Underbid - It will be noted from the responses that the WPC may not be able to
10   process the required number of work items exactly. Instead it may be able to process less, as in case 2.

* Matching bid - In addition the WPC may be able to increase its resources (by hiring extra workers for instance), in order to process the extra work items.

* Overbid - However it may be the case that the WPC is unable to bring in just
15   enough workers to process the extra work items, resulting in an overbid. The underlying reason for this is assumed to be a non-linear constraint between the number of extra workers brought in and the volume of work that they can perform, per unit time.

When an overbid occurs, the CA Agent has two main options, it can
20   either canvass other WPC Agents to see if they would be willing to relinquish work (albeit at a price) which could then be reallocated to the overbidding agent, or the overbid can be accepted and paid for by the CA Agent, which will then have some 'credit' with a particular WPC Agent, which should be taken up as soon as possible.

25   There are two other types of response that a WPC Agent can make:

* Accept - this is an extreme form of the matching bid, as the WPC Agent undertakes to match all the requested attributes (i.e. rate and quality and start time and duration and price);

* Reject - this is an extreme form of the underbid, as the WPC Agent indicates
30   that it is currently unwilling/unable to undertake any change to its existing work rate.

When the CA Agent receives a bid that it finds potentially acceptable, it must put the bidding WPC Agent on hold, whilst it explores the implications of

accepting that bid. It is not computationally feasible to consider every bid received as being potentially acceptable, as the time taken to explore all of the possibilities would be a non-polynomial function of the number of WPC Agents present. (See discussion below of search options for the WMC Agents.)

5

Price Profiles

Referring to Figure 7, in order to perform this calculation, the agents need data relating to the "real" work processing world. The type of data needed can be
10   abstracted from a detailed knowledge of the resources available at the WPC 100 and the capabilities of these resources, and stored as a price profile at each of the WPC Agents. There will be one price profile 700 per work category. An example is shown in Figure 7 and the use of price profiles is illustrated in the section "Scenario Types" below.

15        The x-axis of the price profile represents a delta (a change in the rate of work). A positive value indicates the increase in the rate at which the WPC 100 is being asked to process work. A negative value represents work items (the volume per unit time) that the CA 105 might want to buy back from the WPC. The y-axis has two meanings. When the delta is positive, it represents the minimum
20   acceptable price (paid by the CA) in order for the WPC to process the extra work items. When the delta is negative, the y-axis represents the penalty price4 (paid by the CA) in order to buy back some work items from the WPC. (A penalty price might arise if a WPC 100 agrees to a CA's request to relinquish work.)

It should be noted that the profiles are non-linear. This has an important
25   consequence for the negotiation between the agents. If the profiles were linear and WPC Agents provided responses of the form 2 above, then over a period of time the CA Agent could reason that the relationship was linear and even infer the exact form of the linearity. If this were the case it would no longer need to ask each WPC Agent whether or not it could perform the work items and at what
30   price, since it could determine this itself. However, a more realistic situation is that the profiles are dynamic over time. Thus the exact form of the non-linearity is not fixed.

30

It should also be noted that there could of course be more than three WPC agents. Although there are types of interactions between three agents that have no counterpart in two agents, all of the interactions involving four or more agents have a counterpart (albeit simpler) in three agents. Thus the example

5   described above, using three agents, is sufficient to underpin much richer interactions. (However, it should be noted that the algorithm described below in the section entitled "Scenario Types" will scale non-linearly (i.e. quadratically) with the number of WPC Agents.

10  **Pricing Considerations**

The detailed pricing to be performed by the WPC agents, briefly discussed above, do not form part of the present invention; the mechanism by which a resource is priced is specific to the particular resource concerned.

15  Nonetheless, in general, the process of deriving a price consists in monitoring the availability of resources (e.g. number of staff available); monitoring the current volume of work; calculating the available fraction of resources (the unallocated resources); storing some data representative of the cost of the resources available (staff costs and overheads); and calculating a price for doing

20  new work which takes into account the costs of resources, and in the incremental cost of taking on the new work.

For example, if a given work handling centre consists of four persons, all fully occupied, then the cost of taking on additional work includes the cost of an extra person, which introduces a fixed cost regardless of whether the additional

25  work occupies 1% or 100% of the capacity of that person, as well as variable costs related to the volume of work. Thus, a cost for performing exactly the volume of work specified can be quoted.

By way of contrast, where one person is 50% occupied by existing work, the price of taking on additional work will rise at a (relatively low) first rate until

30  that person is fully occupied, at which point the rate of rise of price with volume of work will change (due to the need for an additional person).

Accordingly, the pricing mechanism is preferably arranged to be capable of locating such discontinuities in the price/volume curve shown in Figure 7 for

amounts lower than the required quantity specified by the CAA agent, and to be capable of bidding for an amount which efficiently utilises the available resources; in this case, by accepting work equivalent to the 50% of one person workload, or, in general, a volume of work falling within a discontinuity in the price/volume

5    curve.

In this way, WPCAs are capable of bidding economical prices for lower volumes of work than those specified.

One example of an algorithm will now be described.

Briefly, referring to Figure 13, in a step 3002, the WPC agent receives a

10    bid quantity signal from the CAA agent. In a step 3004, it calculates a price for supplying the required quantity.

In a step 3006, a plurality of prices for progressively lower quantities (lower than the initial quantity signalled) are also calculated, and in a step 3008, it is determined whether any of these would be more profitable (i.e. enable the

15    provision of services at a lower price). If so, then in a step 3010, the lowest such price and quantity are substituted for the original price and quantity.

In a step 3012, the price and quantity thus derived are signalled back to the CAA agent.

Then, in a step 3014, the WPC agent awaits a further signal from the

20    CAA agent.

If the signal received is a rejection, then the negotiation process is terminated, and all temporarily stored price and quantity data is erased.

Alternatively, a further bid quantity signal may be received, specifying a different quantity. In this case, the previously transmitted quantity and price data

25    is temporarily stored, since the CAA agent may later return to accept (as discussed below) that proposal. Then, ignoring the temporarily stored data, the process returns to step 3002 to calculate a fresh price for the revised quantity signalled (on the assumption that the WPC agent will not receive an acceptance of the previous offer).

30    Finally, the signal received at step 3004 may be an acceptance signal from the CA agent, specifying one of the price/quantity combinations previously indicated by the WPC agent. In this case, the temporarily stored details of that

offer are retrieved by the WPC agent, which proceeds to set up a contract in step 3006, in accordance with Contract Net Protocols, or as discussed above.

It will be understood that where, as discussed above, each work flow process can handle multiple different types of work, the calculation of an optimal

5      quantity and price is performed by varying several of the types of work to attempt to find different local minima of price against volumes.

### Residue Queues

10     The CA Agent maintains two residue queues which are used to record details of short-term excesses and deficits in the number of work items respectively.

### Excess Queue

15

The agents attempt to negotiate a work distribution plan. There are situations when the required number of work items cannot be distributed exactly, leading to a surplus. These surplus items are maintained on an excess queue, on a per work item category basis. Each entry is of the form:

20

*Work Item Category, Excess Rate*

When a new delta arrives, the CA Agent will modify it by combining any excess items with the delta.

25

### Deficit Queue

In the process of distributing the work items whilst attempting to minimise cost, the CA Agent may make an undertaking to an agent to give it

30     some extra rate of work items when it can. This undertaking, or deficit, is recorded on a deficit queue, on a per work item category basis. The entry is of the form:

*Work Item Category, Deficit Rate, Owed to*

When a new delta arrives the CA Agent will first look to see if the deficits
can be reduced or removed altogether by taking work items from the delta, and
5   pre-assigning them to the WPCs that are in deficit.

An example is as follows.   Initial excess and deficit queues are as shown
in Table 5 and Table 6, and an incoming delta is shown in Table 7. Note that in
each case 'rate' refers to a given volume of work per unit time.

| Work Category | Excess rate |
|---------------|-------------|
| A             | 5           |
| B             | 0           |
| C             | 0           |

10

Table 5: Example excess queue

| Work Category | Deficit rate | Owed To |
|---------------|--------------|---------|
| A             | 0            | -       |
| B             | -10          | WPC1    |
| C             | -6           | WPC3    |

Table 6: Example deficit queue

15

| Δidentity = Δ1 Work Category | Source   | Rate | Quality | Start Time | Duration |
|------------------------------|----------|------|---------|------------|----------|
| A                            | External | 0    | 0       | NOW        | INFINITE |
| B                            | External | +50  | 0       | NOW        | INFINITE |
| C                            | External | +30  | 0       | NOW        | INFINITE |

Table 7: Example external delta

Assuming that an external delta arrives (details as per Table 7), then the CA
20   Agent will first remove the deficit owed to WPC1 to get a modified deficit queue

(as per Table 8) and revised delta (as per Table 9). The CA Agent also removes
the deficit owed to WPC3 to give a modified deficit queue and revised delta as
per Table 10 and Table 11 respectively.

| Work Category | Rate | Owed To |
|---|---|---|
| A | 0 | - |
| B | 0 | - |
| C | -6 | WPC3 |

Table 8: Modified deficit queue

| Δidentity = Δ1 Work category | Source | Rate | Quality | Start Time | Duration |
|---|---|---|---|---|---|
| A | External | 0 | 0 | NOW | INFINITE |
| B | External | +40 | 0 | NOW | INFINITE |
| C | External | +30 | 0 | NOW | INFINITE |

Table 9: External delta, after first revision

| Work Category | Rate | Owed To |
|---|---|---|
| A | 0 | - |
| B | 0 | - |
| C | 0 | - |

Table 10: Final version of modified deficit queue

| Δidentity = Δ1 | Source | Rate | Quality | Start Time | Duration |
|---|---|---|---|---|---|
| A | External | 0 | 0 | NOW | INFINITE |
| B | External | +40 | 0 | NOW | INFINITE |
| C | External | +24 | 0 | NOW | INFINITE |

Table 11: External delta, after second revision

Finally, any excess queue items are added to the delta, resulting in a modified delta as per Table 12 and modified excess queue as per Table 13.

| Δidentity = Δ1 | Source | Rate | Quality | Start Time | Duration |
|---|---|---|---|---|---|
| A | External | +5 | 0 | NOW | INFINITE |
| B | External | +40 | 0 | NOW | INFINITE |
| C | External | +24 | 0 | NOW | INFINITE |

5   Table 12: Final version of modified external delta

| Work Category | Excess Rate |
|---|---|
| A | 0 |
| B | 0 |
| C | 0 |

Table 13: Final version of modified excess queue

10      This residue queue management activity results in the formation of a partially complete work distribution plan, as shown in Table 14.

| Work Item Category | WPC1 | WPC2 | WPC3 |
|---|---|---|---|
| A | - | - | - |
| B | 10 | - | - |
| C | - | - | 6 |

Table 14: Partial work distribution plan

15

Scenario types

       Two types of scenario can be demonstrated. The first illustrates what happens when the workload offered to the WMC changes (in terms of either rate

20  or composition) and shows how the agents negotiate with each other in order to formulate a plan for redistributing the resultant work items i.e. an external delta.

The second type shows how the CA Agent reacts when WPC Agents request modifications to existing WPC workloads i.e. an internal delta. Examples of each type of scenario are described below.

5   Steady-state operation

Figure 8 illustrates the initial steady-state conditions. There are a number of SLAs in place between each of the WPCs 100 and the CA 105, which are causing WPCs 1, 2 & 3 to be operating at close to normal capacity, well below

10   normal capacity and at normal capacity, respectively. Each WPC has a different price profile (as indicated by the graph adjacent to each) which determines its bidding policy for work items offered to it by the CA. Note that both of the CA Agent's residue queues are initially empty.

The introduction of an external delta represents a change in the workload

15   offered to the WMC, whereas an internal delta represents a request to modify a WPC's existing workload. (The structure of a delta is discussed above.)

The mechanism for receiving a new external delta into the system is similar to that carried out for an internal delta, the difference being that the 'Source' will be set to the unique identifier of the WPC that raised the exception.

20   Load re-balancing is performed ultimately by the CA, and results in the updating of appropriate SLAs. However, the re-distribution plan that guides the CA is generated as a result of inter-agent negotiation. Negotiation can be initiated by either the CA Agent (for external deltas) or a WPC Agent (internal deltas).

Referring to Figure 9, the first scenario involves an influx of additional

25   work to the WMC, which is seen as an external delta at the CA. In Figure 9 the delta has been shown as a simple, positive change in work rate. Deltas can of course be more complex, involving changes in quality requirements etc.

When an external delta is received by the CA 105, the details are passed to the CA Agent with a request for it to instigate a negotiation. The CA Agent's

30   first step is to look in its deficit queue to see if some or all of the additional work can be accommodated using any spare capacity that has previously been negotiated and paid for but was not utilised for some reason. It will also look in its excess queue to see if any outstanding work is to be added to the incoming delta.

As the demonstration scenario starts from the steady-state this will not be the case first time round so the next step is to see which of the WPCs is willing and able to absorb the extra work.

To initiate this procedure a request message is broadcast to all the WPC
5    Agents asking if any of them is able to handle all or part of the additional work. This message will contain the following information:

* A message ID - a unique identifier for the message;
* A sub-goal ID - a unique identifier relevant to the resolution of one item in a
10   delta;
* A context ID - a unique identifier relevant to each negotiation;
* The details of the delta - as shown in Table 2.

On receipt of the request, each WPC Agent will check to see if its WPC
15   can handle all or part of the additional workload without further negotiation. Depending on available resources, a WPC Agent will be able to bid as follows:

* Matching bid - All of the work regardless of type;
* Overbid - More work than was actually offered;
20   * Underbid - Less than all of the work which will be a combination of one or more of the following:
* all of some types;
* some of all types;
* some of some types.
25   * Accept bid - All of the work as per the requested quality and price parameters;
* Reject bid - None of the work.

If it can handle any additional work, the WPC Agent then consults its price profile to calculate a price for doing the work. This price is then returned to
30   the CA Agent in a bid message which will consist of:

* A message ID;
* The same context ID;

* The same sub-goal ID, if required;

* The price for doing all or part of each type of work as appropriate.


The offers are made on the assumption that they are only valid for the

5  lifetime of the negotiation.

The CA Agent will then wait for a response (or time-out, after a suitable

period) from all interrogated WPC Agents and will then have to decide which offer

to accept, if any. Assuming that the CA Agent is to attempt to get the work done

at the minimum price to itself, it will check if the price supplied will cover all the

10  additional workload. If not, the CA Agent will have to consider the cost of

assigning the unallocated work if it accepts the lowest cost partial offer. To do

this, a message is sent to all WPC Agents (except the one that returned the

lowest offer first time around) requesting prices for accepting the residual

workload. This message will be similar to the first, except that it also includes a

15  maximum price which the CA Agent will have calculated by considering the

difference between the lowest offer (regardless of work accepted) and the offer

for doing all of the work, if applicable.

The next message will be of the form:


20  * A message ID;

* The same sub-goal ID, if required;

* The same context ID;

* The amount of work of each type that is required;

* The maximum total price for doing the work.

25

Again, the WPC Agents may respond with a price for doing all or some of

the work as detailed above. If there is no offer for doing all the work then the

offer that would cover the most workload should be considered, and this process

loop is repeated until the delta is satisfied or the closest approximation is reached.

30  The aim of the CA Agent may be to get the work done leaving the lowest

possible deficit residue, in this case the strategy for selecting offers would be

slightly different in that the percentage of work distributed would be the most

important criteria to consider rather than lowest price.

Consideration should be given here to the situations where there is no close match or no match that would cover all of the work.

Assuming the minimum-cost scenario, a run-through of the stages of negotiation and calculation is shown below:

5      As shown in Figure 10, the details of the external delta arriving at the CA Agent are:

(20A, 10B, 5C)

A request to do the work is sent out to all WPC Agents

10      (20A, 10B, 5C)?

Offers are returned by the WPC Agents

(x1A, x2B, x3C) @ U from WPC1 Agent

(y1A, y2B, y3C) @ V from WPC2 Agent

15      (z1A, z2B, z3C) @ W from WPC3 Agent

The cost factor is calculated:

U . | (20 - x1)A + (10 - x2)B + (5 - x3)C | from WPC1 Agent

V . | (20 - y1)A + (10 - y2)B + (5 - y3)C | from WPC2 Agent - all the

20      work but not the cheapest

W . | (20 - z1)A + (10 - z2)B + (5 - z3)C | from WPC3 Agent

Assuming y1 = 20, y2 = 10 and y3 = 5 (an exact match from WPC2 Agent but not at the lowest price) the next narrowcast messages (to all except WPC2 Agent)

25      are:

| ( (20 - x1)A + (10 - x2)B + (5- x3)C ) | < = | (V - U) |

| ( (20 - z1)A + (10 - z2)B + (5- z3)C | ) < = | (V - W) |

where the other WPC Agents are requested to bid for the difference between the WPC2 Agent bid (an exact match but not the lowest price) and the WPC1 Agent

30      bid (the cheapest price but not an exact match ).

Assume the reply from WPC2 Agent to the first narrowcast message gives a deficit in work for the right price then a narrowcast message is sent out (to WPC3 Agent in this case) for a bid that covers the difference in work required

and work tendered for so far. From all this received information the CA Agent can generate the total cost to itself for placing the work.

Referring to Figure 9, the second scenario will involve one of the WPC Agents requesting that its workload be changed. This would occur if, for example,

5    the WPC Agent was notified that its WPC's resources had been reduced in some way. It is envisaged that when the CA Agent receives such a request from a WPC Agent, it deals with it in a similar way to if it had received an external delta, i.e. a process similar to the one described above is enacted. The one difference is that the first broadcast message is not sent to all WPC Agents as above but is sent to

10   all WPC Agents EXCEPT the one that raised the internal delta in the first place.

WPC2 Agent generates a request to reduce its type B workload:

(OA, -5B, OC)

15   The CA Agent reverses the polarity of any non-zero rates, checks its residue queues and sends a narrowcast message to the other two WPC Agents:

(OA, 5B, OC)?

As before the WPC Agents return bids for the additional work and the process

20   continues as before:

(OA, 5B, OC) @ X from WPC1 Agent

(OA, 5B, OC) @ Y from WPC3 Agent

If the returned bids are not acceptable then the contents of the delta are added to

25   the residue queue for allocation at a later date.

(As before, the delta has been simplified for clarity both in the text and in Figure 9).

**Preferred Search Options**

30

Search algorithms are employed to find one or more candidate solutions to the problem of re-distributing work categories. The CA builds its distribution

plan to satisfy a given delta by inviting bids from the WPCs and evaluating the responses.

The strategy implemented is for the CA to initially broadcast an invitation to bid for the entire delta, but subsequent invitations (to bid) contain a partial
5  delta, and are narrowcast.

The choice of which partial deltas to narrowcast to which WPCs will influence the speed at which a solution is reached, and the relative quality of that solution. It is preferred that these choices be guided at each step, by examining the responses received and proceeding with the 'best' outcome (however that is
10  defined). Any WPC that contributed to the best outcome is removed from the list of WPCs to narrowcast the next partial delta to. This procedure is repeated until one of the following conditions is met:

a) All WPCs are eliminated;

b) There is no residue of work;

15  c) The search times out.


Even with guided choice, there are a number of decisions to be made:.

* How is the 'best' outcome determined at each step?

* How many solutions should be generated?

20  * How should any residues be handled?


It is proposed that the responses from the WPC Agents are evaluated as follows:

i)  For each work category, calculate the difference between the quantity
25  requested, and the quantity offered;

ii) Multiply each difference term by the offered price for that work category;

iii) Multiply each difference term by the defect coefficient (derived from the quality measure) for that category;

iv) Multiply each difference term by the CA's (inverse) priority coefficient for that
30  work category;

v) Sum all the modified difference terms, and take the modulus of the result.

This single term gives the 'cost' of each response, as seen by the CA. The best response is the one with the lowest cost (with an exact match having cost = zero), and it is this response that is used as the basis of the next partial delta.

5   The overall cost of a particular solution is given by the sum of the costs of all the accepted responses (N.B. price is just one of a number of contributing factors), up to the point that condition a), b) or c) is met.

Many other search strategies are possible, but this one is bound to terminate, and will produce solutions in quadratic, rather than factorial, time. In
10  case of identical results some policy rule could be implemented to select the most appropriate.

Whenever terminal condition a) is encountered, there may be a residue of work to deal with. Although this is assumed to be positive (i.e. some work remains unplaced), there are conditions in which a negative residue may exist (i.e.
15  the CA has committed to placing more work than is currently available for distribution). This situation could arise, for example, where a WPC would be obliged to employ extra resources in the short term to meet the additional demand, but cannot exactly match the demand. It chooses to bid for more work than was offered to it (as it deems the financial cost of the overbid to be
20  marginal).

In either case, the CA Agent is responsible for recording any residues, and attempting to eradicate them as part of the process of handling subsequent deltas that it receives.

Again, choices arise - this time over how the CA manages these residues.
25  It can record them automatically, or it can seek confirmation from the User. Once recorded, subsequent management of the residues can either be automatic, or on an advisory basis. Automatic management strategies can be tempered by setting thresholds, e.g. the User could be advised when the total number of excess/deficit items exceeds X% of the WMC's 'normal' work capacity.

## SUMMARY

Embodiments of the present invention show several innovative aspects. These include for instance:

5   the use of at least two categories of agent, one category providing a quasi-centralised control of workload distribution;

the capability of inter-agent negotiations to relate to multiple work-types simultaneously;

management of over and under bidding by the use of deficit queues, or some like

10   record, for subsequent allocation of the relevant workload and/or, potentially, for subsequent renegotiation of available resources; and

distribution of work according to class or type rather than by individual work items;

response of the agent-based system in substantially real time to incoming data

15   from the real environment, rather than via a planning and scheduling capability.

Thus, to summarise, the present invention uses a multistage signalling process to obtain quotes for different combinations of the different resources.

This might seem simply to duplicate calculation and use valuable computing and signalling resources

20   In fact it would do so if the relations between volume of work and price for the different resources were linear. In this case, a single program could simply calculate all possible variations of workflow.

However, in many situations they are far from linear and accordingly, by providing two layers of agents to negotiate, with the second layer (the WPCAs)

25   aware of the actual status and pricing at their respective resources, non-linear behaviour can be handled.

Additionally, by providing multistage signalling, in which different combinations of quantity and price from different agents are modelled, allocations which use these nonlinearities in the supply/price curves for each resource can be

30   tested and the best selected.

Finally, by maintaining a baseline solution, higher cost allocations can speedily be rejected.

The invention has been described for allocation of varying workflows (i.e. rates of work) and is particularly useful for that since the long term agreements in place reduce the volume of signalling required to occasions where the rate changes rather than to each occasion when new work arrives.

5     However, it can also in principle be used for allocation of batches of work, tasks or resources rather than rates of doing them.

It will be apparent that the invention is equally suitable to, for example, the supply of electricity from multiple power stations each having a different price/power (i.e. rate of doing work) capacity; in this case, a WPCA may be

10    provided at each generator station and a CA agent at the central electricity generating headquarters for a region of a country.

It will be understood that the above described process for performing several different rounds of negotiation involves the CA agent being able to put the WPC agent "on hold", and to evaluate further bids from that agent without

15    abandoning the possibility of accepting an earlier offer from that agent. Whilst this could be applicable in trading environments (such as stocks or securities dealing) it is particularly suitable for agent-controlled resource allocation in which all agents form part of a single system, or (more generally) the CA agents are not in competition for the services of the WPC agents, so that the WPC agents have

20    no economic reason for not going "on hold" (as they might have worth a servicing competing CA agents). Accordingly, such uses of the invention are particularly preferred.

Although the above described embodiments match a need for resources (for work to be done) to a plurality of resource providers, it will be apparent that

25    they could also be used to match a resource supply to a plurality of resource users in a manner clear to the skilled reader hereof.

It will be understood that the invention is not limited to the foregoing embodiments but encompasses variations and modification thereto which would be apparent to the skilled person. Protection is hereby sought for any and all

30    novel subject matter, and combinations thereof.

**CLAIMS**

1.      A resource allocation system comprising computing means for executing:

at least one first computer program module for determining allocation

5    matching one or more of a plurality of resource supplies to a resource

requirement;

at least two second computer program modules for monitoring the state

of a corresponding one of said resource supplies and for communicating with said

first program module.

10

2.      A system according to claim 1, in which said resources comprise means

for performing a task at a plurality of different rates over time.


3.      A system according to claim 2, in which said first computer program

15   module and said second modules are arranged to store records relating to rates at

which respective said resource supplies will perform tasks, and said resources are

allocated in accordance with said agreements until changes between the rate at

which tasks need to be performed and the sum of the rates stored in the

agreements occur.

20

4.      A system according to claim 1, for allocating an original quantity, in

which the first module is arranged to:

a)      signal, to at least two said second modules, a first request for resource

supply indicating a required quantity;

25   b)      receive at least one response, each specifying a resource supply quantity,

from respective ones of said second modules;

c)      select one from said responses;

d)      add the supply quantity specified therein to a running quantity total; and

e)      if said running quantity total is less than said desired quantity;

30   f)      signal, to those of said second modules other than that last selected in

step c), a request for resource supply indicating the difference between said

original amount and said running total; and

g)      repeat steps c) to g).

5.     A system according to any preceding claim, in which the first module is arranged to determine a first possible allocation between resources corresponding to respective second modules, and then to determine at least one second possible allocation, and to choose one of said possible allocations.

6.     A system according to claim 5 when appended to claim 4, in which the first allocation is an allocation to the resource, if any, for which the amount signalled by its respective second module exactly matches the original amount.

7.     A system according to any preceding claim, in which each second module is arranged to signal a price to said first, and said first is arranged to determine said allocation based on at least one factor including said price.

8.     A system according to claim 7 when appended to claim 5, in which the first module is arranged to;

a)     store a baseline price derived from that signalled by a first of said second modules, said first being that which was selected for said first possible allocation;

b)     select another of said second modules which responded at step b) of claim 4;

c)     perform steps d) to g) of claim 4, to derive said at least one second possible allocation.

9.     A system according to claim 8, in which said first module is arranged to; form a running price total corresponding to the sum of the prices signalled in those responses for which the quantities form the running quantity total.

10.     A system according to claim 9, in which said first module is arranged to; replace said baseline price with a running price total for a second possible allocation which is lower than said baseline price.

11.     A system according to claim 9, in which said first module is arranged to;

cease to derive a second possible allocation when the corresponding said running price total exceeds said baseline price.

12. A system according to any of claims 8 to 11, in which said baseline price
5 is derived jointly from the price signalled by a said second module and at least one other factor relating to the resource with which that second module is associated.

13. A system according to any preceding claim, in which said second modules are arranged to signal to said first to indicate a variation on their resource
10 supply capacity.

14. A system according to claim 13 in which said first computer program is arranged, in response, to varying the allocation.

15 15. The system of claim 14 when appended to claim 4 in which the first module is arranged to perform the process of claim 4 to vary said allocation.

16. A system according to any preceding claim, wherein said resources comprise telecommunications channels.
20
17. A system according to any preceding claim, wherein said modules comprise software agents.

18. A method of resource allocation comprising the steps of:
25 signalling a resource requirement quantity to a plurality of resource suppliers;
receiving replies from the resource suppliers which specify different quantities and prices;
temporarily selecting a first of said replies;
30 temporarily selecting a second of said replies from a second of said resource suppliers;
signalling to resource suppliers other than said second for supply of the different between said second and said original quantity; and
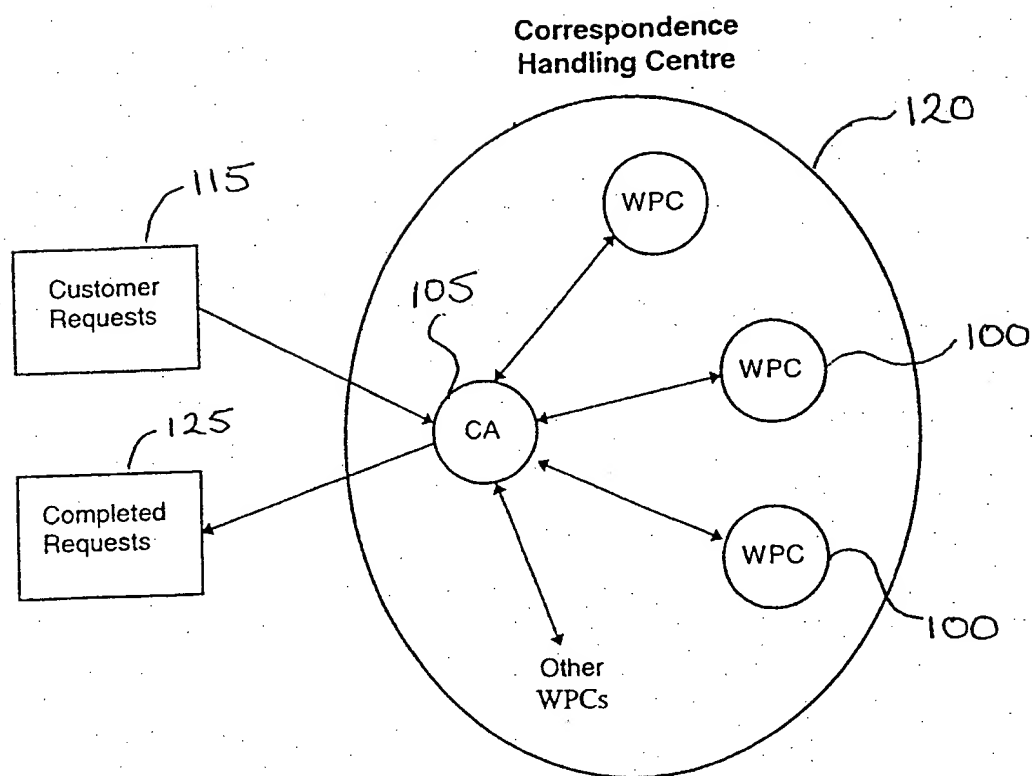
selecting one of said temporary selections as said resource allocation.

19.    A workload management system for use in controlling resource allocation for tasks amongst a plurality of resources for performing the tasks, the system
5    comprising at least one computer controlled by set of software modules, the set being provided with:

i)    at least one workload input, for receiving data relating to workload;

ii)    at least one resource-data input, for receiving resource-formulated data
10    relating to resource availability;

iii)    communication means for inter-module communication;

iv)    data storage means which stores, in use:

a)    data representing at least one set of conditions for task performance together with identifiers for at least two of the software modules;

15    b)    data representing an allocation of said tasks amongst said plurality of resources for performing the tasks; and

c)    data representing a collaborative protocol for determining collaboration between the software modules

wherein one or more administrative software modules, selected from the
20    set of software modules, is provided with said workload input, a plurality of resource-related software modules from the set are each provided with a resource-data input, and the administrative software module(s) control(s) workload management amongst said resource-related software modules.
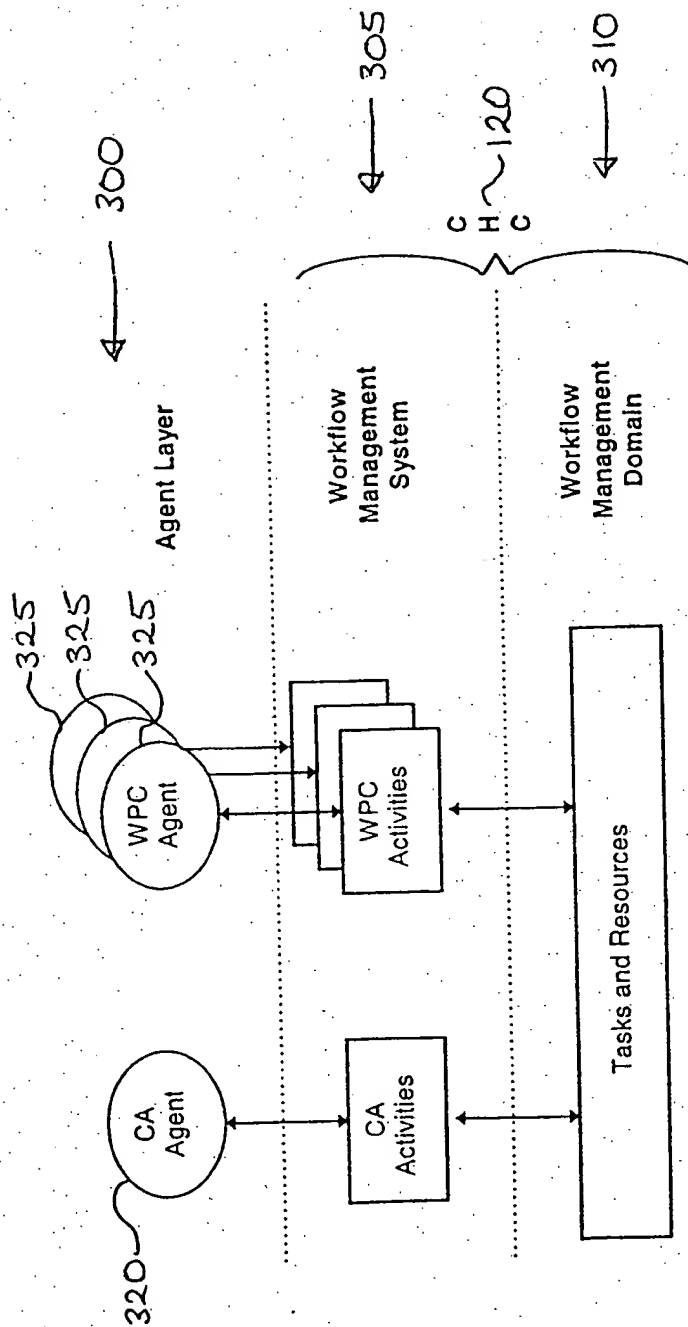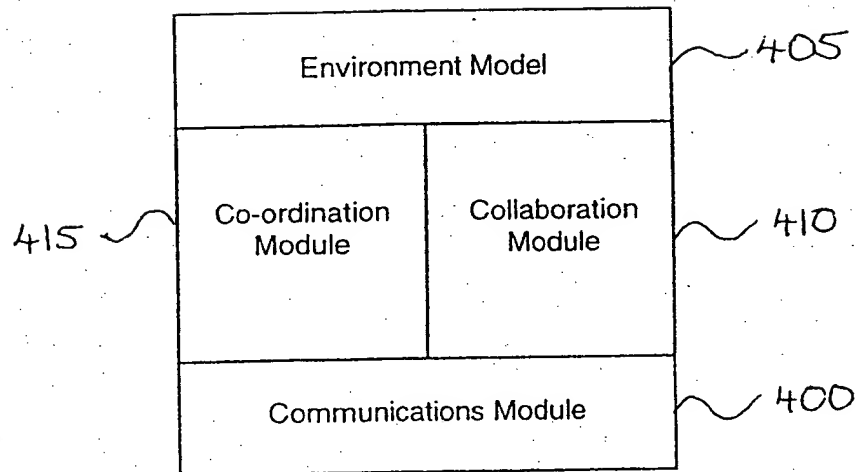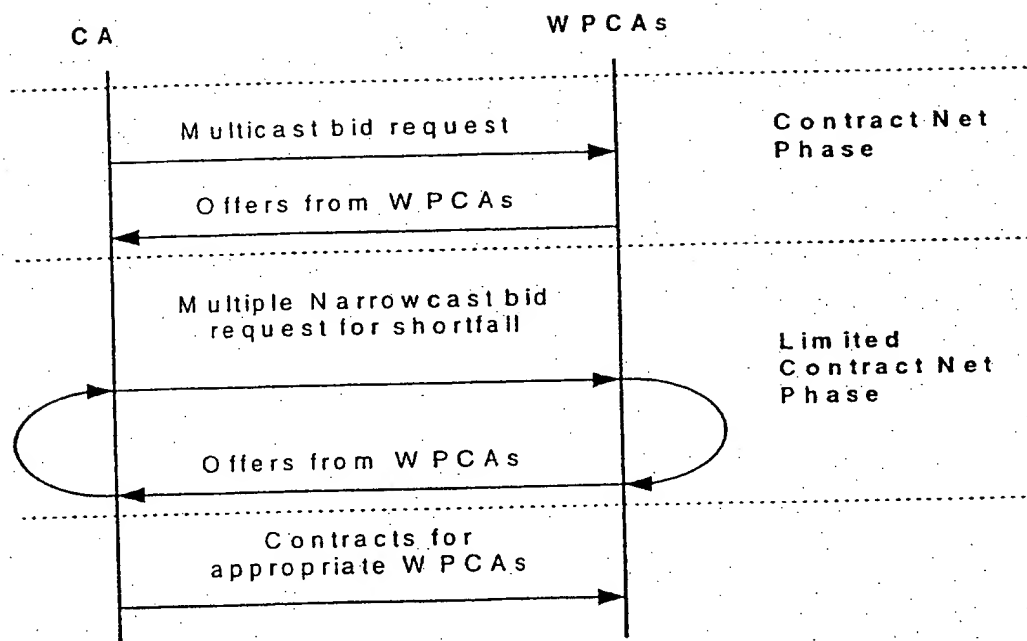
**FIG. 1**

2/15

## FIG. 2



Customer Requests — 115

Initial activities performed by the CA — 200
- Reception — 230
- Classification — 225
- Distribution — 205

Processing WPC 1 — 210
Processing WPC 2 — 210
Processing WPC N — 210

Final activities performed by the CA — 220
- Inspection — 215
- Dispatch

Completed Requests — 125
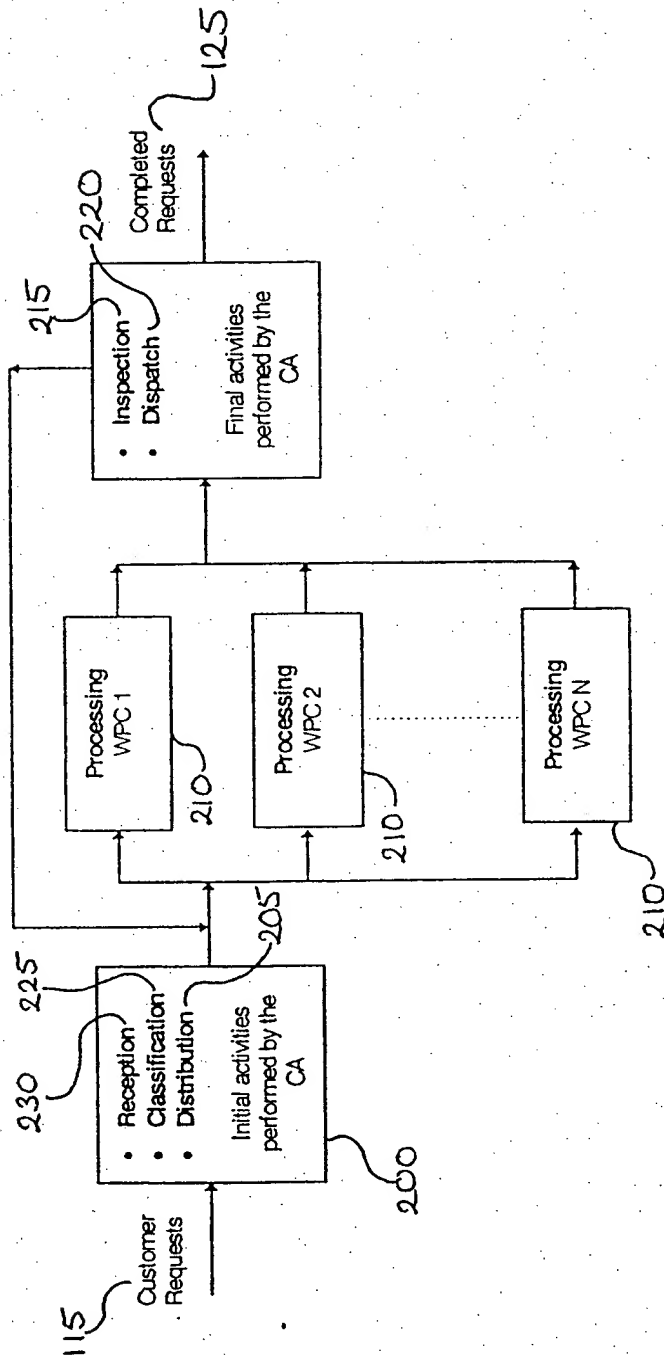
FIG. 3

FIG. 4

## FIG. 5

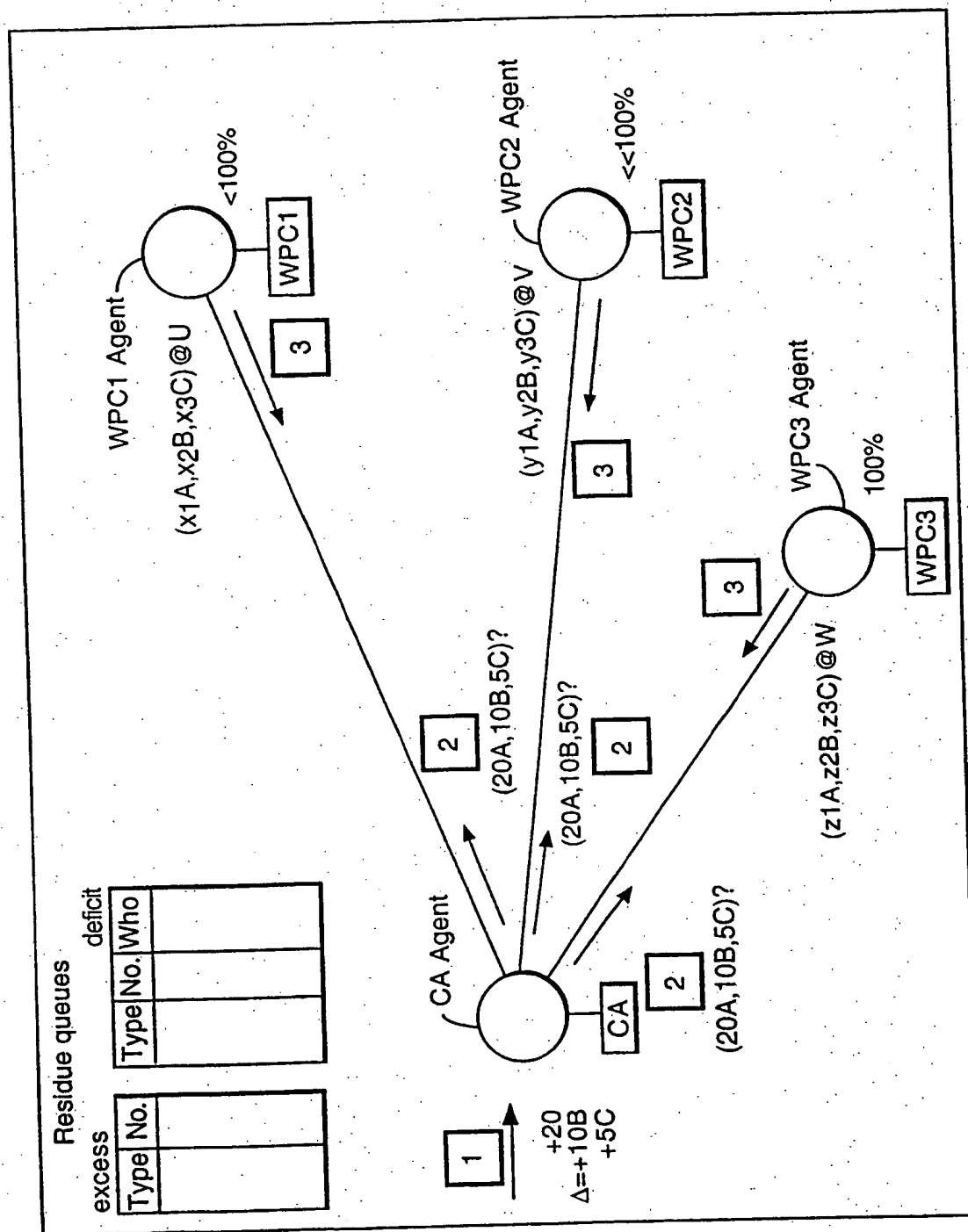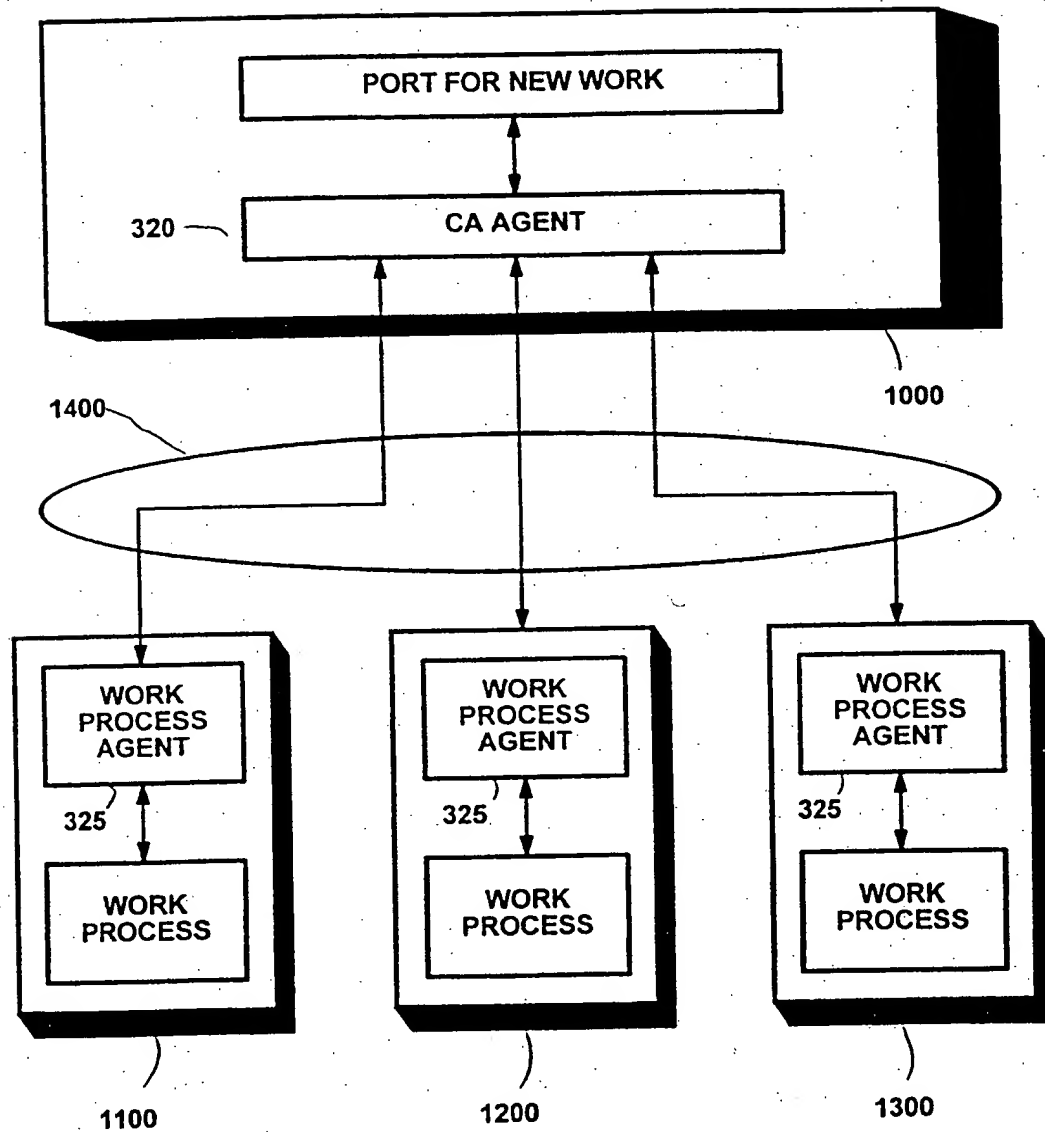| | Workflow Agents | | CHC Model |
| --- | --- | --- | --- |
| | CA Agent | WPC Agent | Workflow Management System |
| Frontware (Clients) | Control GUI-JAVA | Control GUI - JAVA | Workflow Admin. Tools |
| Middleware (Agent-WfMs) | Workflow ORB - CORBA 2 complaint C++ server with JAVA clients | | |
| Middleware (Agent-Agent) | Agent Communications ORB - JAVA | | |
| Backware (Servers) | Workflow Agent Architecture - JAVA | Workflow Agent Architecture - JAVA | Proprietary Workflow product |
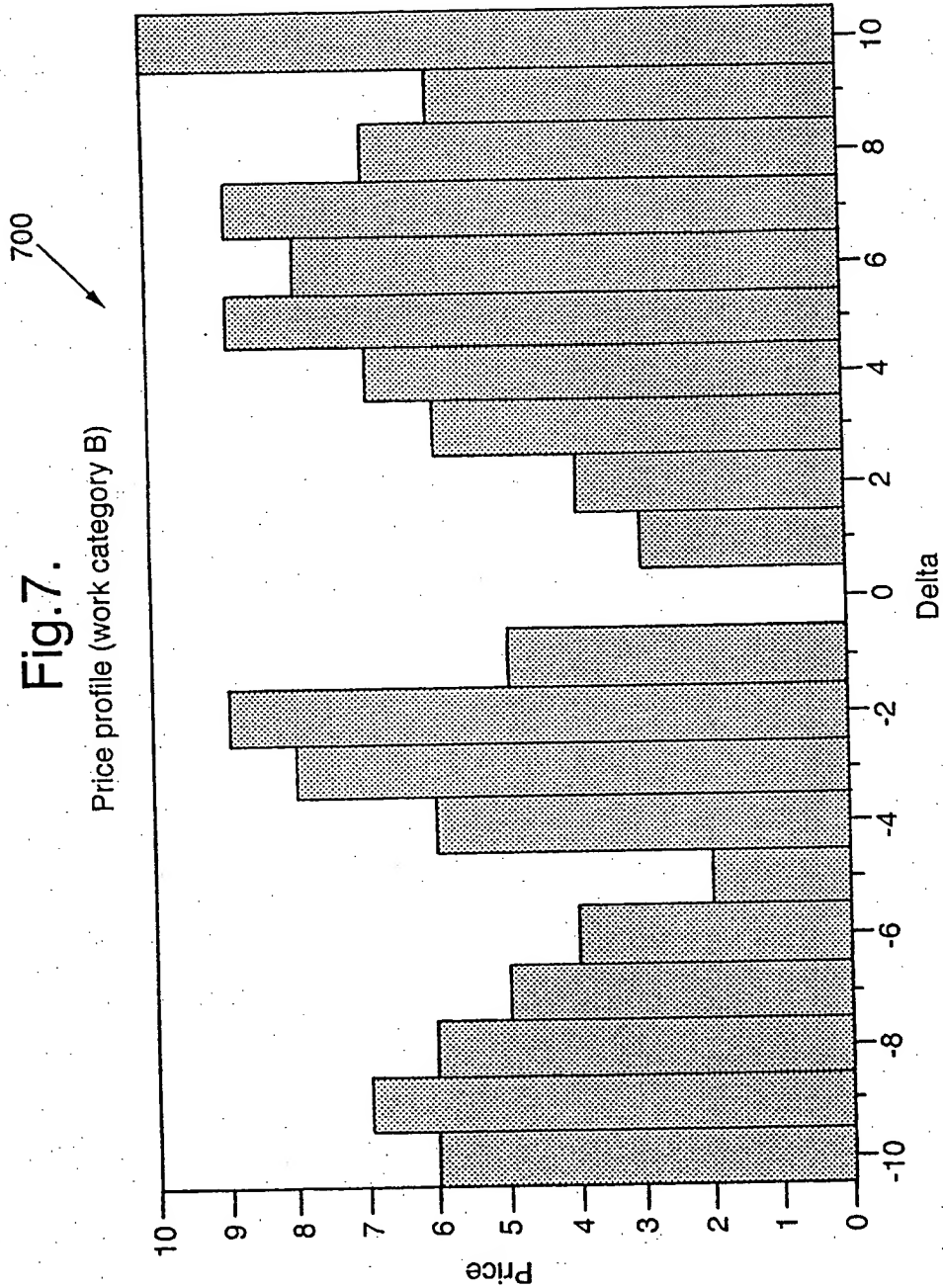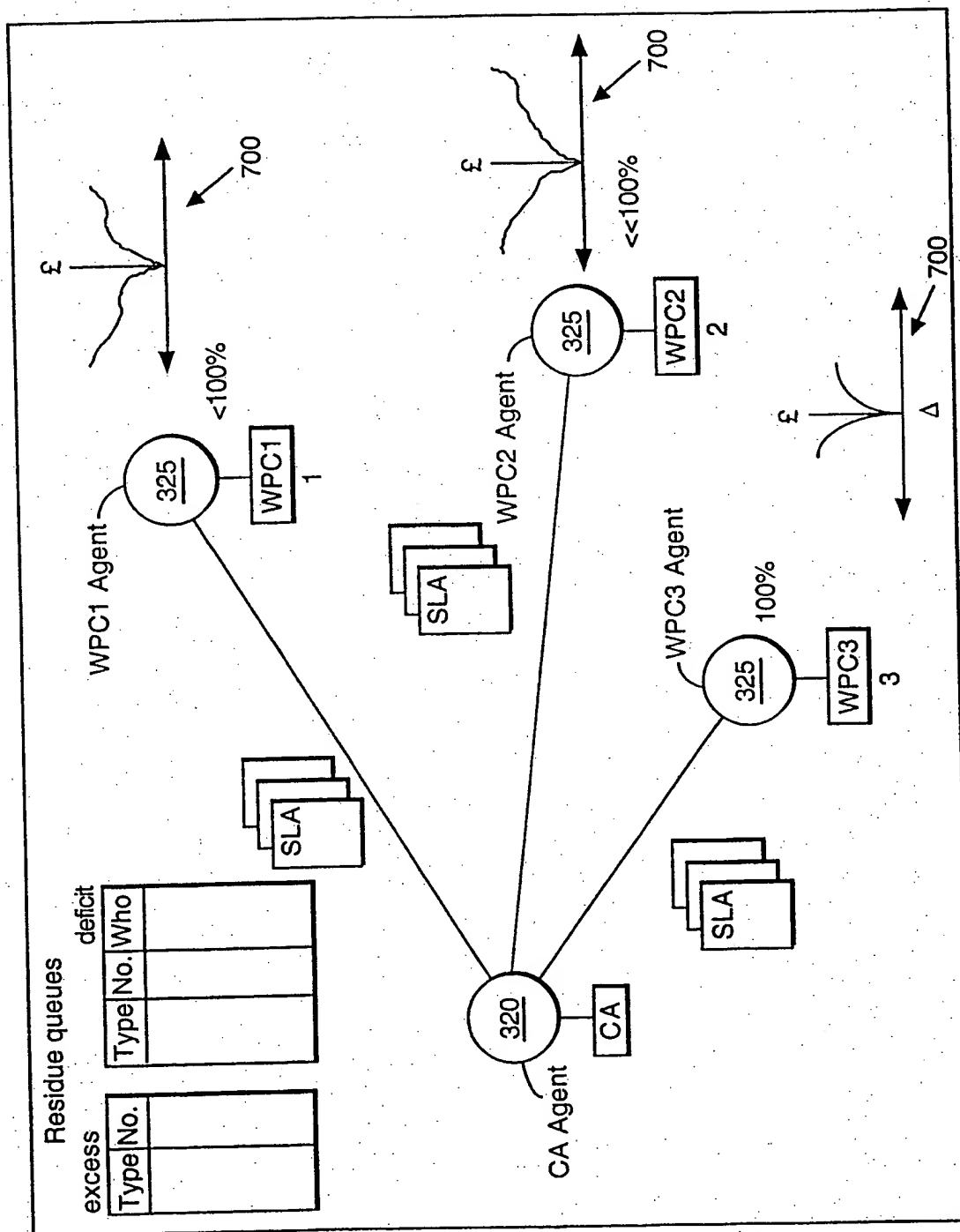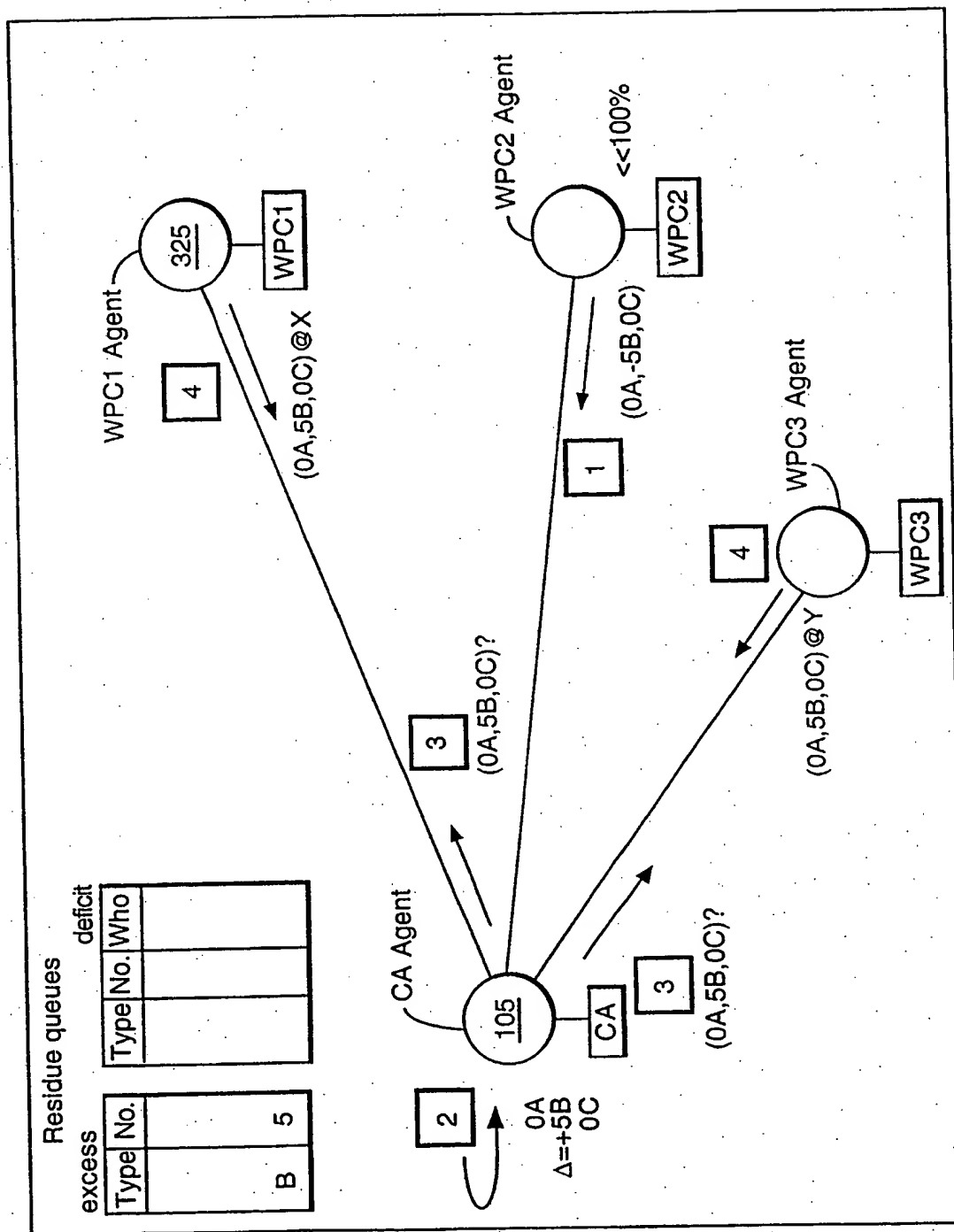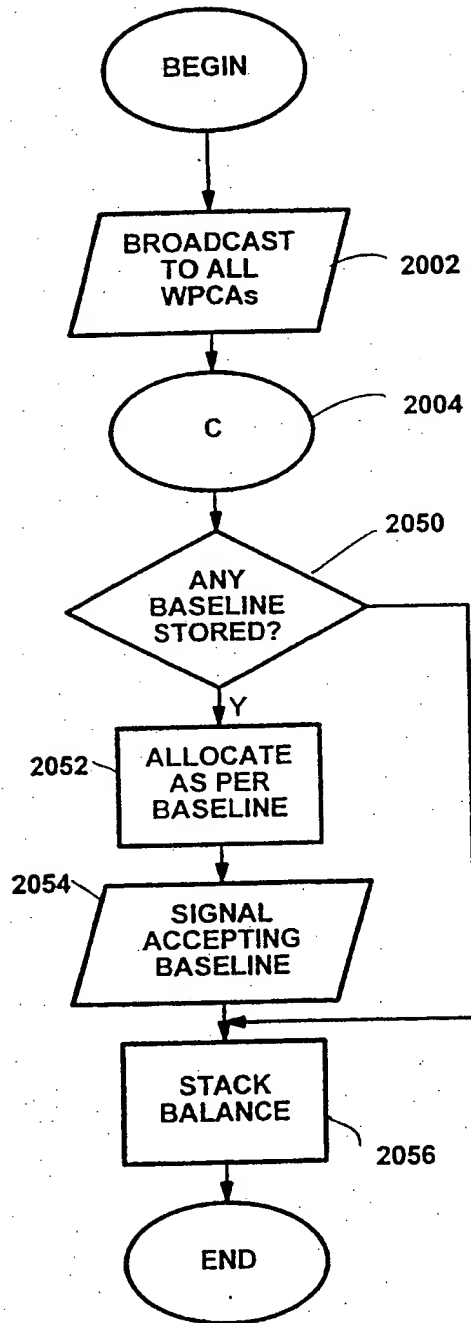
# FIG. 6

Fig.7.

Price profile (work category B)

700

Fig.8.

Fig.9.

Fig.10.

FIG. 11

BEGIN

BROADCAST
TO ALL
WPCAs — 2002

C — 2004

ANY
BASELINE
STORED? — 2050

Y

2052 — ALLOCATE
AS PER
BASELINE

2054 — SIGNAL
ACCEPTING
BASELINE

STACK
BALANCE — 2056

END

**FIG. 12a**

C

RECIEVE
REPLIES — 2006

A — 2008

RANK
BY COST — 2030

Y

LAST
? — 2032

N

SELECT
LOWEST
REMAINING — 2034

B — 2036

RETURN

**FIG. 12b**

**FIG. 12c**

2040 — CALCULATE REMAINING QUANTITY

2042 — NARROWCAST TO OTHER WPCAs

2044 — C

2046 — RETURN
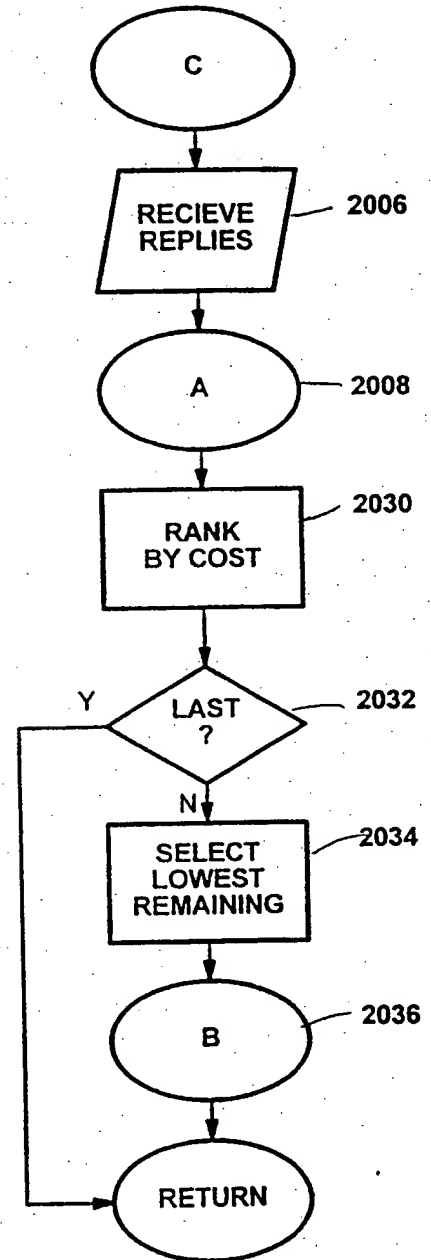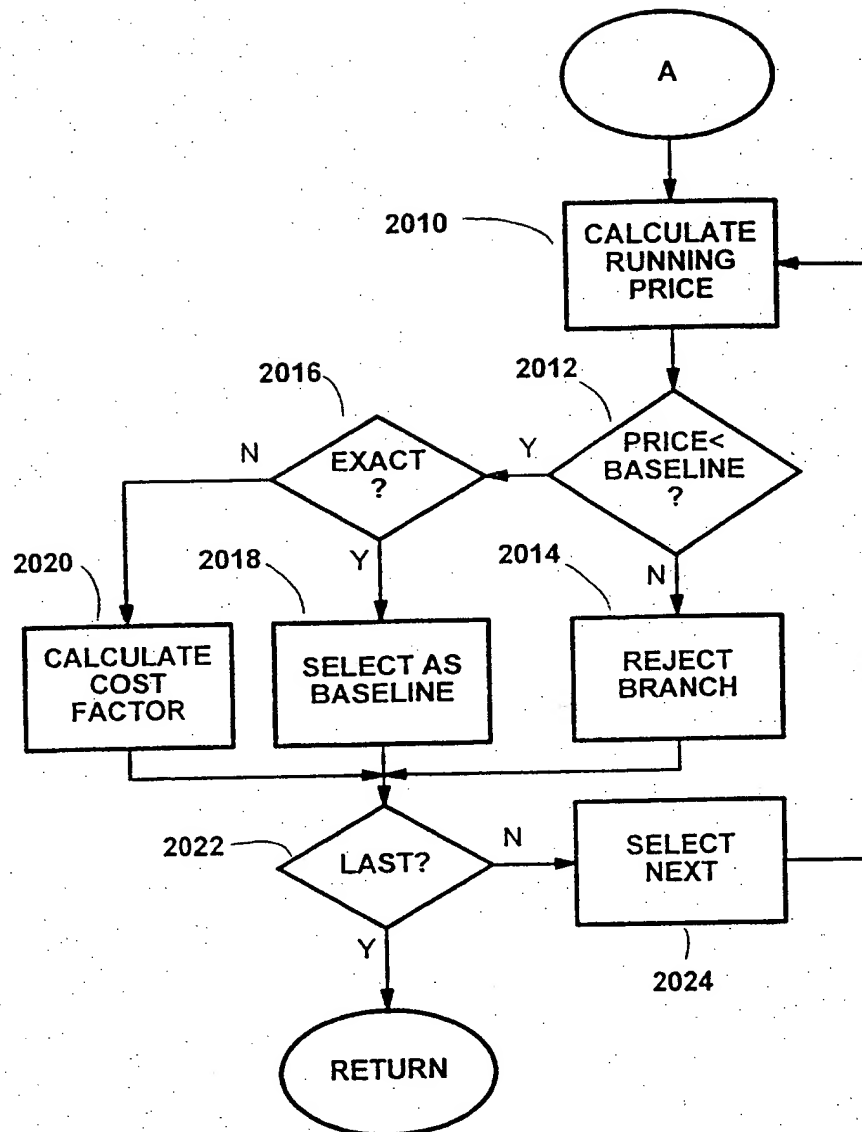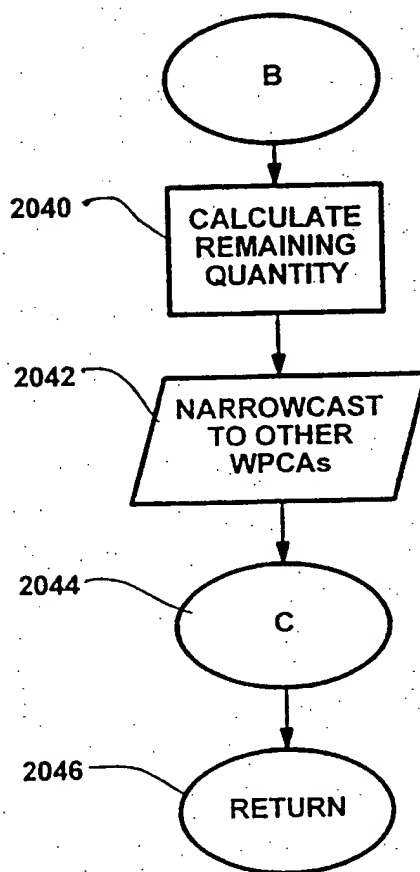
## FIG. 12d
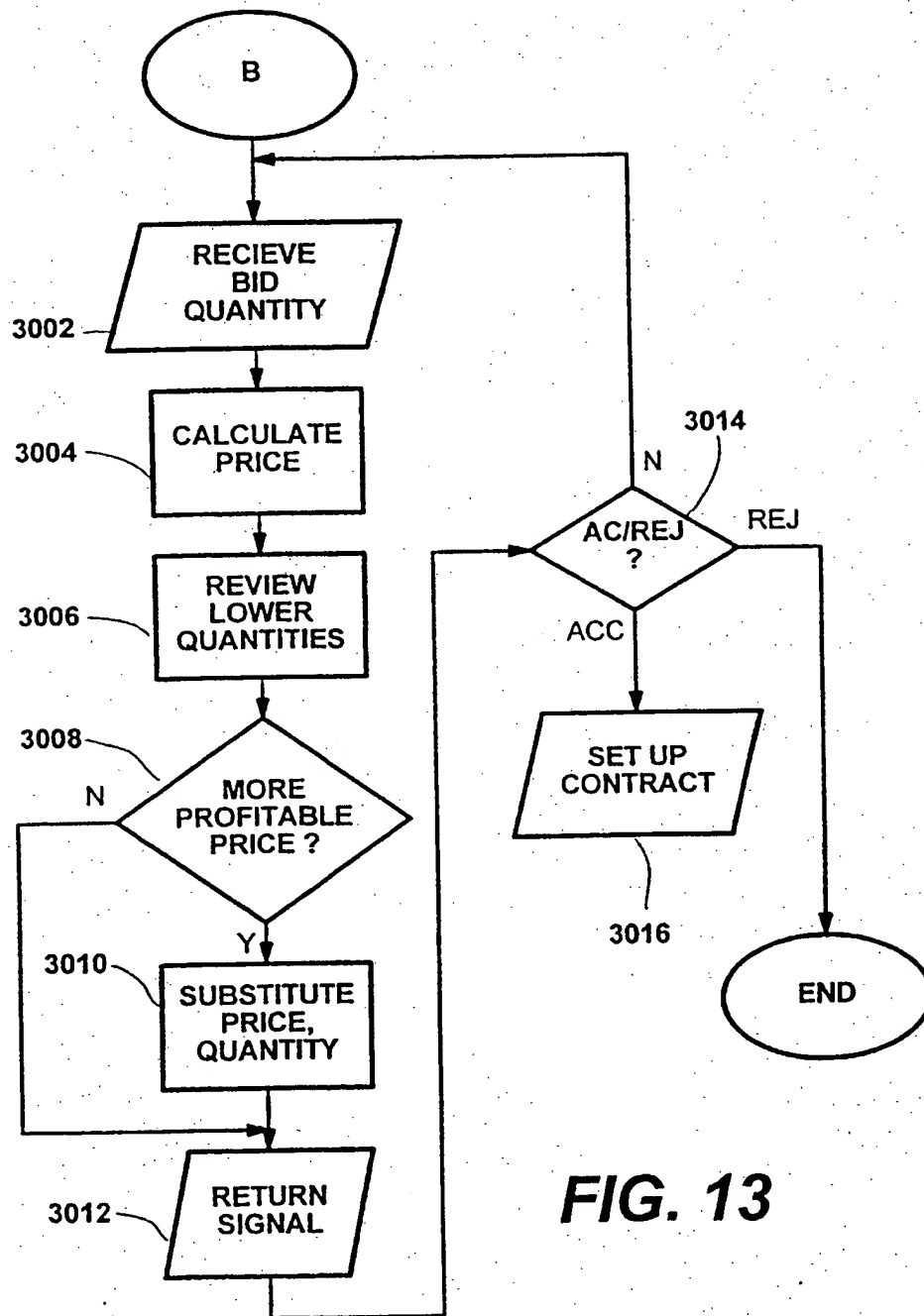
**FIG. 13**

# INTERNATIONAL SEARCH REPORT

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 6    G06F9/46       G06F17/60

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6    G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | CHAVEZ A ET AL: "Challenger: a multi-agent system for distributed resource allocation" PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS, PROCEEDINGS OF 1ST INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS, MARINA DEL REY, CA, USA, 5-8 FEB. 1997, pages 323-331, XP002092534 ISBN 0-89791-877-0, 1997, New York, NY, USA, ACM, USA | 1,17 |
| A | see page 325, left-hand column, line 16 – page 326, left-hand column, line 4 --- -/-- | 18 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 8 February 1999 | 18/02/1999 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016 | Michel, T |

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | RAMOS C: "A holonic approach for task scheduling in manufacturing systems" PROCEEDINGS. 1996 IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION (CAT. NO.96CH35857), PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, MINNEAPOLIS, MN, USA, 22-28 APRIL 1996, pages 2511-2516 vol.3, XP000772583 ISBN 0-7803-2988-0, 1996, New York, NY, USA, IEEE, USA | 1,17 |
| A | see page 0 see paragraph 3.3 see paragraph 4 | 4,18 |
| A | SANDHOLM T: "Limitations of the Vickrey auction in computational multiagent systems" ICMAS-96 PROCEEDINGS. SECOND INTERNATIONAL CONFERENCE ON MULTI-AGENT SYSTEMS, PROCEEDINGS OF ICMAS '96: 2ND INTERNATIONAL CONFERENCE ON MULTIAGENT SYSTEMS, KYOTO, JAPAN, 10-13 DEC. 1996, pages 299-306, XP002092535 1996, Menlo Park, CA, USA, AAAI Press, USA see the whole document | 1,18 |
| X | GOSCINSKI A ET AL: "RESOURCE MANAGEMENT IN LARGE DISTRIBUTED SYSTEMS" OPERATING SYSTEMS REVIEW (SIGOPS), vol. 24, no. 4, 1 October 1990, pages 7-25, XP000273166 see paragraph 2 see paragraph 3; figure 5 | 19 |
| A | GB 2 194 086 A (HITACHI LTD) 24 February 1988 see the whole document | 19 |
| A | NWANA H S ET AL: "AN INTRODUCTION TO AGENT TECHNOLOGY" BT TECHNOLOGY JOURNAL, vol. 14, no. 4, October 1996, pages 55-67, XP000635333 see the whole document | 1,19 |
| A | WO 96 03692 A (BRITISH TELECOMM ;BUTT JOHN (GB); IRELAND PAUL STUART (GB)) 8 February 1996 see the whole document | 19 |

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

1

# INTERNATIONAL SEARCH REPORT

Intern    ial Application No

PCT/GB 98/02944

| Patent document cited in search report | | Publication date | Patent family member(s) | | Publication date |
|---|---|---|---|---|---|
| GB 2194086 | A | 24-02-1988 | JP | 2533495 B | 11-09-1996 |
| | | | JP | 63034056 A | 13-02-1988 |
| | | | US | 4852001 A | 25-07-1989 |
| WO 9603692 | A | 08-02-1996 | AU | 689925 B | 09-04-1998 |
| | | | AU | 2986695 A | 22-02-1996 |
| | | | CA | 2194766 A | 08-02-1996 |
| | | | EP | 0772825 A | 14-05-1997 |
| | | | FI | 970288 A | 24-01-1997 |
| | | | JP | 10503306 T | 24-03-1998 |
| | | | NO | 970315 A | 24-01-1997 |
| | | | NZ | 289894 A | 24-11-1997 |